Education
OHIO STATE UNIVERSITY
AUG 28 1959
LIBRARY

Journal of PROJECTIVE TECHNIQUES

PUBLISHED BY THE SOCIETY FOR PROJECTIVE TECHNIQUES AND RORSCHACH INSTITUTE, INC.

TABLE OF CONTENTS

Symposium: Current Aspects of the Problem of Validity
Suggestions for the Delineation of Validational Studies
EDWIN S. SHNEIDMAN 259
Variability of Responses and the Stability of Scores and Interpretations of Projective Protocols
DONALD W. FISKE 263
Structured and Projective Tests: Some Common Problems in Validation ..... PAUL E. MEEHL 268
A Study in Validation of the Thematic Apperception Test ..... WILLIAM E. HENRY and JANE FARLEY 273
What is a Criterion? ..... EVELYN HOOKER 278
Validity and Methodology in Projective Tests
NORMAN L. FARBEROW 282
Problems in the Validation of Projective Techniques
KENNETH R. LITTLE 287
Further Check on the Diagnostic Validity of the Ames Danger Signals ..... LOUISE BATES AMES 291
Sex Identification at Different Ages Using the Draw-A-Person Test ..... R. L. BUTLER and F. L. MARCUSE 299
The Effect of Examiner Status Upon Rorschach Performance
FRANCES A. CAMPBELL and PAUL R. FIDDLEMAN 303
The Perceptual Organization TAT Score: Number, Order, and Frequency of Components ..... RICHARD H. DANA 307
The Relationship Between Rorschach Responses and Attitudes Toward Parents ..... CYNTHIA ENGEL 311
Types of Pre-Cognitive Defensive Organization in a Tachistoscopic Experiment ..... ULF KRAGH 315
Projective Test Assessment of the Effects of Sleep Deprivation
NATHENE TURK LOVELAND and MARGARET THALER SINGER 328
The Structured Doll Play Test as a Projective Technique for Use With Children ..... DAVID B. LYNN and ROSALIE LYNN 335
Relationship Between Story Length and Situational Validity of the TAT ..... CAROLINE TAYLOR MacBRAYER 345
The Blacky Pictures as Personality Measures for Undergraduate Areas of Specialization ..... MAX G. MAGNUSSEN 351
Sexual Identification in Responses to Rorschach Card III
MARVEN O. NELSON, WILLIAM WOLFSON and RALPH LO CASCIO 354
The Experimental Induction of the Defense Mechanism of Projection ..... ROBERT ROSENTHAL 357
The Integration of Determinants and Content in Rorschach Interpretation ..... DAVID SHAPIRO 365
Book Reviews 374
General Newsletter 375
Announcements 376

Volume 23 X September 1959 No. 3

OHIO STATE UNIVERSITY EDUCATION LIBRARY

OFFICERS OF THE SOCIETY FOR PROJECTIVE TECHNIQUES AND RORSCHACH INSTITUTE, Inc.

President ..... WILLIAM E. HENRY
President-Elect ..... GORDON F. DERNER
Past President ..... LEOPOLD BELLAK
Representative at Large
Eastern ..... RUTH L. MUNROE
Western ..... EDWIN S. SHNEIDMAN
Secretary ..... HELEN H. DAVIDSON
Treasurer ..... HARRY V. MCNEILL
Editor ..... BRUNO KLOPFER
Executive Editor ..... BERTRAM R. FORER

• • •

ADVISORY COMMITTEE

FRANKLIN G. EBAUGH
LAWRENCE K. FRANK
BRUNO KLOPFER, *Chairman*
NOLAN D. C. LEWIS
MORTIMER M. MEYER
GARDNER MURPHY
DAVID VORHAUS

JOURNAL OF PROJECTIVE TECHNIQUES, published quarterly by the Society for Projective Techniques and Rorschach Institute, Inc. Annual subscription (calendar year) $8.00, foreign $8.50. Single copies $2.00. Publication office: 210 E. Wilson Ave., Glendale 6, Calif. Subscriptions, changes of address and business matters should be addressed to the Society for Projective Techniques and Rorschach Institute, Inc., 210 E. Wilson Ave., Glendale 6, Calif., or Adelphi College, Garden City, N. Y. Address all Editorial Matters to: Bertram R. Forer, 2170 Live Oak Drive, E., Los Angeles 28, Calif. Entered as 2nd class matter at Glendale, Calif.

Copyright 1959 by the Society for Projective Techniques and Rorschach Institute, Inc.

JOURNAL OF PROJECTIVE TECHNIQUES

VOL. 23 SEPTEMBER, 1959 NO. 3

• • •

EDITOR

BRUNO KLOPFER

EXECUTIVE EDITOR

BERTRAM R. FORER

ASSOCIATE EXECUTIVE EDITORS

WALTHER JOEL
JEANNE M. REITZELL
EDWIN S. SHNEIDMAN

STAFF EDITORS

EVELYN CRUMPTON
FLORENCE DIAMOND
NORMAN L. FARBEROW
HERMAN FEIFEL
FRED J. GOLDSTEIN
JAMES HOLT
NICHOLAS ROSE
LEONARD B. OLINGER
JOHN WEIR
IRLA LEE ZIMMERMAN

CONSULTING EDITORS

JOHN E. BELL
EUGENIA HANFMANN
ROBERT R. HOLT
LOIS B. MURPHY
MARIA A. RICKERS-OVSIANKINA
L. JOSEPH STONE

Symposium: Current Aspects of the Problem of Validity[1]

Suggestions For the Delineation of Validational Studies

EDWIN S. SHNEIDMAN
V. A. Center, Los Angeles

This symposium on "Current Aspects of the Problem of Validity" is sponsored both by the Division of Clinical Psychology and the Society for Projective Techniques. Our topic relates to one of the most timely issues that might have been selected by either group. It is one that is important to our practice and fundamental to our science. As chairman, my own contribution to this symposium is one that might be called "A Fugue and Variations on a Four-Part Invention by Cron-bach," and is essentially a suggested modification of the types of validity formulated by Cronbach's APA Committee on Test Standards in 1952.

One hardly needs to reiterate that the APA Committee on Test Standards proposed four types of validity, as follows:

1. "Predictive validity, which denotes correlation between test measures and *subsequent* criterion measures.

2. "Status validity, which denotes correlation between the test and *concurrent* external criteria.

3. "Content validity, which refers to the case in which the specific type of behavior called for by the test is the goal of teaching or some similar activity, as in an academic achievement test.

4. "Congruent validity which is established when the investigator demonstrates what psychological attribute a test measures by showing correspondence between scores on a test and other indications of the state or attribute. This type of validity is used for tests intended to measure a construct arising from some theory; the validation consists of evidence that the scores vary (from person to person or occasion to occasion) as the theory would imply. Essentially, in congruent validity the meaning assigned to test scores is substantiated by demonstrating that scores are consistent with deductions from the theory from which the meaning derived. This validation process is much the same as that involved in evaluating a theory itself."

My own suggested modification is essentially a spelling out of the several dimensions involved in any validational study. I think that I might go so far as to propose that each of these be indicated at the beginning of any validational study. The headings I suggest are as follows:

1. Validity for *whom,* in which (a) the subjects, (b) the examiners, and (c) the judges would be described. It makes a lot of difference, for example, whether the 700 Rorschachs being reported in a normative matching study were given to Columbia sophomores or Columbian peasants; whether the examiner was an ABEPP grandfather or his comely young granddaughter, and whether the judges were tyros or Tyroleans. This notion is, of course, a sort of tautology in that it begs the question of what are the relevant dimensions in terms of which the subjects, examiners, and judges need to be described.

2. Validity for *when,* in which one (or more) of the three logical temporal positions would be indicated: past, present, and future. Correlations with future criteria are, of course, called *prediction,* and with past criteria are called *postdiction.* For correlations with more-or-less present criteria (such as "present status") I

[1] Symposium presented at the American Psychological Association Convention, Washington, D. C., August 30, 1958.

have suggested the term *paridiction,* from the word "equal".

3. Validity for *which.* Under this rubric one should indicate whether correlations are with essentially *factual* variables or with essentially *conceptual* variables. This category again begs the question because it is precisely in this area that most of our problems concerning validity reside. There are, in clinical psychology, relatively few "facts", if we define a fact as a perception about which several intelligent lay people would instantaneously concur. When we view a body lying in a pool of blood and see a holographic document with suicidal content, it is a "fact" (Alfred Hitchcock aside) that the individual has committed suicide; but in a not-yet-dead individual "being suicidal" is almost entirely conceptual. Most of what we deal with in clinical psychology are not facts but hypothesis-tinged data.

4. Validity for *what.* Under this heading it is suggested that (a) the nature of the stimuli be delineated, and more, that (b) the nature of the continua or dimensions along which the stimuli are believed to be spread be delineated. Dr. Farberow will in his paper expand on this notion and indicate some of the theoretical and practical implications of developing specific psychological testing techniques which focus on the elaboration of stimuli along a single psychological dimension.

5. *How* Validity is established. This would be an indication of the basic statistical procedures involved. What I have in mind is the notion that it is not only what the statistics choose to tell, but is also in part which statistical procedures were chosen. Clinically this means that even the statistician must be aware of his own trends toward latent homoscedasticity.

But it is to the notion of congruent validity of tests, and the idea that the validational process is much the same as that involved in evaluating a theory itself, that I wish to return for the remainder of my introductory remarks, and make the to-me-obvious, but I fear in some circles somewhat iconoclastic comment that our psychological tests or projective techniques are not validatable until the personality theories that give them their intellectual permissiveness themselves gain greater construct validity.

And that is not all. The matter is further complicated inasmuch as the data on the other side of the equation, the so-called criterion measures, are themselves extremely complex and little understood. Let me give two examples, using homicide and schizophrenia—I had promised that I would not dwell upon either suicide or the MAPS test during my presentation. On the first thought it would seem impressive if one were to have not ten, but 200 Rorschach records of individuals who had committed homicide and records of 200 individuals who had not. But a moment's reflection will lead one to the common-sense consideration that individuals kill other individuals in a variety of ways (in a liquor-store hold-up, in the lover's bedroom, on the highway, on the field of battle, etc.) —and that *each* of these occurs with a variety of motives and mixed feelings. Putting all the "homicidal" records together could only yield a kind of psychometric goulash. (Incidentally, suicide is even more complicated.) And, now let us, for a few moments, consider schizophrenia—at first blush a perfectly wonderful external criterion for status validity. Whose schizophrenia do we mean? How shall we define schizophrenia?

1. Is schizophrenia as Kraepelin believed, a mental disease process resulting from an auto-intoxication following a disorder of metabolism produced by a disordered secretion of the sex glands?

2. Or is it, from Arieti, a primitive form of rationality, based on pathological reasoning in terms of attributes of the predicate?

3. Or, from Alexander, is it a state

in which the ego loses its synthetic function of harmonizing the different instinctual demands?

4. Or, from Jung, does the peculiarity of this disease lie in the startling emergence of an archaic psychology accounting for the innumerable points of contact with mythological material, so that what we take for original and individual creations are mostly products which can only be compared with those of antiquity?

Incidentally, Jung states that there are two types of schizophrenia (asthenic and spastic), Beck says there are six, and I had been taught all along that there are really four.

5. Or, is schizophrenia, as Meyer felt, to be thought of as a habit disorganization resulting from a progressive maladaption with increasing use of substitute reaction, instead of effective ones.

6. Or, do we really understand schizophrenic delusions as a projection of repressed passive homosexual desires? I think the formula is as follows: I love him, I do not love him, I hate him, No he hates me. Now I have it, I hate him because he hates me. We hate each other, I don't think that I will invite *him* to my next gay party.

All these are concepts. But there are, happily, three solid facts about schizophrenia. The only three known facts about schizophrenia are:

1. We do not know exactly how to cure it; its remediation is unknown.

2. We do not know exactly what causes it; its etiology is unknown—except when we say that it is cryptogenic.

3. We do not know its nature. We do not know what it is, if it is an it, which it probably ain't.

In 1958, we in clinical psychology (and individuals no less in psychiatry) have to deal primarily with the methological problems attendant to *construct* validity.

Our task is in some way similar to measuring a floating cloud with a rubber band—in a shifting wind. I used to hear that projective techniques were like taking a plug out of a watermelon. This is a delusion of agronomy.

The science of clinical psychology relates primarily to hypotheses, not to facts. Whether one believes in Freud's, Adler's, Jung's, Murray's, Allport's, Kelly's, or any other individual's personality theory, one still has to subsume his cerebrations under the purview of an essentially untested hypothesis, nothing more. As a clinician, an individual may (in terms of loyalty) be from Vienna, or Topeka, or Baltimore; but as a scientist he must be from Missouri.

Nor do I believe that the physical sciences should be our models. My notion is that we do not do well to attempt to emulate physicists (simply because they have more observable facts interlaced among their hypotheses) but rather.I suggest that we look *within* our own field, that we reexamine (with a view toward escaping the tactical and strategic errors of theory-building) the intellectual maneuvers of Wundt, Titchener, Dewey, Watson, MacDougall, etc., not to go back to them, but better to see that they were only trying systematically to explore the implications of *their* hypotheses — all their "facts" followed after. Our task might well be to plan our scientific maneuvers so that even if we adhere more or less rigidly to our current hypotheses, at least learning from *our own* past, we might avoid making some of the methodological errors which beset these earlier movements.

It appears that the kinds of validity discussed by the APA committee and by others are all kinds of "research" or scientific validity, whereas it may be that we may have to deal for a while with faith validity or clinical validity wherein we are not concerned with the veridical value of a hypothesis but rather with its professional sense-appeal within the scope of our contemporary conceptual zeitgeist. But on the other hand, it seems that our notions *ought* to be capable of being

metricized so that we can test whether or not the phenomena we think we see, occur beyond chance expectation and with statistical significance.

In order to end on a positive note, I want to say that as a meaningful measure of the nature of the unconscious nothing beats our current projective techniques, whatever the unconscious is—if you see what I mean.

Received December 8, 1958.

Variability of Responses and the Stability of Scores and Interpretations of Projective Protocols

DONALD W. FISKE
University of Chicago

The literature on projective tests contains very little information on the consistency of responses. Very little is known about the stability of responses to the same stimuli at different times and almost nothing is known about the stability of interpretations of such different protocols.

This paper is concerned primarily with these problems. We shall see that responses to the same stimuli tend to be different on different occasions and that such instability of response seems to have serious effects on the stability of interpretations. On the other hand, some scores or ratings derived from successive protocols show fair stability.

From work on the variability of a person's behavior over time, I have come to the postulate that behavior in the absence of external stimulation is completely variable. The more structured the stimulation and the more intense the stimulation (up to a point), the more consistent the behavior. Stimuli exercise constraints on behavior (7).

This position is strongly supported by several kinds of evidence. First let us consider dreams as an example of "behavior" of a sort, behavior in the absence of stimulation.

In the systematic work of Kleitman, Dement, Wolpert, and Kamiya (4), all the dreams a person has during a night are recorded immediately after they occur. For a number of subjects, the dreams of several nights have been recorded. Inspection of these shows no exact repetitions. The content of each dream is essentially unique although one can, of course, find instances of similar themes or the development of a theme, especially in contiguous dreams. Much of the repeating material is associated with the experimental situation[1]. Thus one subject had several dreams, usually the first each night, about the situation, the electrodes, etc. Another had two dreams about people slanting or biasing the dream material they reported. It is worth noting that the general topic of these reported dreams is common across subjects and is largely determined by the general experimental situation.

Another systematic collection of protocols in the relative absence of stimulation is in the work of Rechtschaffen and Mednick (20) on the autokinetic word technique, in which it is suggested that the fixed point of light will trace letters, words, or sentences. Here also, there is no exact repetition of content.[2] Reports from stimulus deprivation studies seem to be consistent with these data. In a study by Goldberger and Holt (10), subjects' verbalizations during an eight hour period of marked stimulus reduction were recorded. Inspection of these protocols, and of post-session interview reports of experiences during the experimental period showed minimal repetition of content.[3] There were two types of exceptions: two subjects repeated the same tune twice; some subjects, especially those who were anxious to escape the situation, repeated content such as wondering what time it was and wondering how he looked as he lay there. The latter type of repeated material related to the situation is strikingly consistent

---

[1] I am indebted to Drs. William Dement and Joe Kamiya for access to these dream protocols.

[2] I am indebted to Dr. Allan Rechtschaffen for access to these protocols.

[3] I am indebted to Drs. Robert R. Holt and L. Goldberger for access to these protocols.

with the repeated dream material noted above.

When relatively unstructured stimuli are present, behavior is more consistent than in the absence of stimulation, but it is still highly variable. One implication of this viewpoint is that a person has a very wide distribution of potential responses, the width or range of the distribution being greater when the stimuli are less constraining. Since projective techniques are rather unstructured, the stimuli and instructions impose relatively little constraint on responses.

Thus Bert Kaplan and his colleagues have found that on a second testing, normal people can give almost as many *new* responses to the Rorschach cards as they gave the first time (28). In a study by Willis (29), subjects gave an average of four or five different plots (themes) to a TAT card in 10 minutes. I have found that subjects can give several responses to each sentence completion stem.

While this evidence is not definitive, I am sure most of you would agree with the notion that a projective stimulus at any point in time for a single person has several potential responses, with varying response strengths. Now what is the evidence on the variability of emitted responses over time?

For word association, Willis and I have some preliminary data indicating that almost one-third to one-half of the responses are changed when the same stimulus word is repeated in a single session or in a second session. (We asked subjects to give two associations to each stimulus. Fifty to 85 per cent of the second associations are changed.)

For sentence completion responses, three studies show that 64% to 80% of the responses are changed from one trial to the next (8, 18, 19). In one study, subjects averaged six or seven different responses over nine trials.

The TAT does not permit similar comparisons. However, I have examined some sets of protocols, each set consisting of three trials 1 to 30 months apart. Rather than estimate similarity, I counted the number of times that the three stories the person told to a card were quite different, one from the other, in manifest content. This occurred in about 20% of the subject-card combinations. In the remaining 80%, the similarity ranged from two stories having one detail in common, which did not appear in the third, to sets of three almost identical stories. Some of the repetition was obviously straight recall, as one subject made explicit who "saw" each time in the blank card a polar bear in a snowstorm!

Swift (24) reports that about 50% of Rorschach responses by 4 and 5 year old children are repeated 2 to 4 weeks later. Siipola, Kuhns, and Taylor (22) indicate that for presentations of parts of these blots, about half the responses were changed on retest. Of these changed responses, about half were similar on the two occasions.

On the Blacky Test, Granick and Scheflen (11) have reported that judges considered that about 57% of the thematic content of two repetitions was similar, but a comparison with a matched group yielded one-third similarity between responses from different children!

For the Picture Arrangement Test, Tomkins and Miner report that 33% of the order responses were changed over a three week interval and 45% were changed over three years (26, p. 37).

The figures we have given are overestimates of response stability because there is much evidence that subjects remember many of their projective responses over several weeks (cf. Swift, 24).

The reader may be thinking that while the manifest content may be changed, the psychological meaning would be essentially the same. To check this, Van Buskirk estimated the changes in psychological meaning over

the nine trials of the Sentence Completion data mentioned earlier. He found a high correlation (.80) between the extent of change in the psychological meaning of a person's responses and the extent of change in the manifest responses. But this was merely suggestive. So we compared the interpretations of different protocols from the same individual (9).

We asked this specific question: are the differences between the interpretations of protocols from the same person less than the differences between interpretations of protocols of different persons? Thus inter-individual variation, variation between individuals, provides a frame of reference for evaluating the extent of intra-individual variability, variation or inconsistency within a person over time.

We found that in 25% of the judge-protocol combinations, the interpretation was more similar to interpretations of protocols from different subjects than to interpretations of other protocols from the same subject. This 25% misclassification was not due to the homogeneity of the subjects. The manifest content differentiated the subjects better than the judges' interpretations.

Once again, such a single study is not definitive. It should be repeated with other tests and with complete batteries. It does, however, suggest that the changes in manifest content on at least one projective test make the interpretations clearly unstable over time.

I hope that the material I have presented has indicated to some of you that the existence of considerable variability over time in projective responses presents a serious difficulty for clinical diagnosis and for personality assessment as a whole. But the picture is not entirely negative: there is considerable evidence that some characteristics of responses are reasonably stable over time.

For example, for the Incomplete Sentence Blank, Arnold and Walter (1) found the Rotter Adjustment Score to have a stability over a week of .82. Churchhill and Crandall (3) report coefficients of .70 to .38 for 6 month to 3 year intervals. Rohde (21) found stability coefficients of .68 and .64 (over an eight month period) for need variables rated on her Sentence Completion Test. The cathections of objects were even more stable: less than 4% were changed. For the TAT, Tomkins (25, p. 6) reports r's from .50 to .80 over two to ten months for ratings using a need-press schema. For a two-month interval, Lindzey and Herman (16) report 17 r's ranging from .00 to .94. For a similar period, Helen Koch[4] has obtained a comparable range of coefficients on the CAT. In the Granick and Scheflen study (11) of the Blacky cited earlier, 75% of the subjects were consistent in their liking or disliking of each Blacky card. For the Sargent Insight Test, Fassett (6) reports several coefficients all below .46. Rorschach scores have stabilities from .98 to −.16 (5, 12, 13, 24). In some unpublished Draw-a-Person test data of Shanan[4] the correlation between total area used for two figures over a six month interval is only .44. This is similar to the only correlations reported in Swensen's review of the literature on the test (23). The evidence for the stability of characteristics of such drawings is not impressive, but Wagner and Schubert (27) cited by Swenson (23, p. 435), report a coefficient of .86 for the stability of ratings of overall quality.

Many of these coefficients indicate that the scores involved are sufficiently stable over time to permit comparisons between groups. However, few are high enough for dependable use in differentiating individuals. There is a possibility that certain temperamental and expressive attributes have higher stability: for example, productivity as measured by word-count. Rechtschaffen[4] has obtained an r of .72 for the autokinetic procedure and Koch[4] found an r of .78 for the CAT.

[4] Personal communication.

But such characteristics are not unique to projective techniques and are certainly not the kind of variable which such techniques are designed to assess.

What does the total picture look like?

1. The actual content of responses to projective tests varies considerably from trial to trial, and there are large individual differences in the extent of this change in content.
2. The stability of *some* scores and ratings based on projective protocols is reasonably adequate.
3. There is reason to believe that changes of responses may have serious effects on the stability of interpretations of total protocols.

The problem is *not* that successive protocols are psychologically inconsistent but that they often complement each other; material in one may not be present in the next. The problem then becomes that of determining the significance of such material for the enduring personality.

One promising approach has been neglected, at least in research literature. This is Murray's suggestion (17, p. 604-5) of a unity-thema: a central theme; compounded from dominant needs and press, which can be seen to run through a person's life.

We need to investigate this notion. Does material from different sources —such as TAT and other thematic protocols indicate the same central theme? And do protocols gathered at different times embody the same thema?

We can safely assume that the "personality" we wish to assess and measure is that part of the person's dispositions and response tendencies which is stable, at least over weeks if not years. Therefore it is essential that we develop methods to distinguish between stable and temporary dispositions, between pre-eminent and subordinate response tendencies. I have found few explicit discussions of this problem (Rohde (21, p. 109); Lindzey (15); Bolgar (2); Klopfer and Spiegelman (14) and no development of adequate methods. It is my conviction that the validity of projective scores and interpretations will continue to be unsatisfactory until such methods are available.

References

1. Arnold, F. C., and Walter, V. A. The relationship between a self- and other-reference completion test. *J. counseling Psychol.*, 1957, *4*, 65-70.

2. Bolgar, Hedda. A re-evaluation of projective theory. In Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., and Holt, R. R. *Developments in Rorschach Technique*. Volume II. Yonkers-on-Hudson, World Book Co., 1956.

3. Churchill, Ruth, and Crandall, V. J. The reliability and validity of the Rotter Incomplete Sentences Test. *J. consult. Psychol.*, 1955, *19*, 345-350.

4. Dement, W., and Wolpert, E. A. Relationships in the manifest content of dreams occurring on the same night. *J. nerv. ment. Dis.*, 1958, *126*, 568-578.

5. Eichler, R. M. A comparison of the Rorschach and Behn-Rorschach inkblot tests. *J. consult. Psychol.*, 1951, *15*, 185-189.

6. Fassett, K. K. A preliminary investigation of the Sargent Test. *J. clin. Psychol.*, 1948, *4*, 45-55.

7. Fiske, D. W. The constraints on intra-individual variability in test responses. *Educ. psychol. Measmt.*, 1957, *17*, 317-337.

8. Fiske, D. W. An intensive study of variability scores. *Educ. psychol. Measmt.*, 1957, *17*, 453-465.

9. Fiske, D. W., and Van Buskirk, C. The stability of interpretations of sentence completion tests. *J. consult. Psychol.*, 1959, *23*, 177-180.

10. Goldberger, L., and Holt, R. A. Experimental interferences with reality contact (perceptual isolation): I. Method and group results. *J. nerv. ment. Dis.*, 1958, in press.

11. Granick, S., and Scheflen, Norma A. Approaches to reliability of projective tests with special reference to the Blacky Pictures Test. *J. consult. Psychol.*, 1958, *22*, 137-141.

12. Holzberg, J. D., and Wexler, M. The predictability of schizophrenic performance on the Rorschach test. *J. consult. Psychol.*, 1950, *14*, 395-399.

13. Kerr, M. Temperamental difference in twins. *Brit. J. Psychol.*, 1936, *27*, 51-59.

14. Klopfer, B., and Spiegelman, M. Methodological research problems. In Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., and Holt, R. R. *Developments in Rorschach Technique*, Vol. II. Yonkers-on-Hudson, World Book Co., 1958.

15. Lindzey, G. Thematic Apperception Test: interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, *49*, 1-21.

16. Lindzey, G., and Herman, P. S. Thematic Apperception Test: A note on reliability and situational validity. *J. proj. Techniques*, 1955, *19*, 36-42.

17. Murray, H. A. *Explorations in Personality*. New York: Oxford, 1938.

18. Osterweil, J., and Fiske, D. W. Intra-individual variability in sentence completion responses. *J. abnorm. soc. Psychol.*, 1956, *52*, 195-199.

19. Raine, W. J., and Hills, J. R. Measuring intra-individual variability within one testing. *J. abnorm. soc. Psychol.*, 1959, *58*, 264-266.

20. Rechtschaffen, A., and Mednick, S. A. The autokinetic word technique. *J. abnorm. soc. Psychol.*, 1955, *51*, 346.

21. Rohde, Amanda R. *The Sentence Completion Method*. N. Y.: Ronald, 1957.

22. Siipola, Elsa, Kuhns, Florence, and Taylor, Vivian. Measurement of the individual's reaction to color in ink blots. *J. Pers.*, 1950, *19*, 153-171.

23. Swenson, C. H., Jr. Empirical evaluations of human figure drawings. *Psychol. Bull.*, 1957, *54*, 431-466.

24. Swift, Joan W. Reliability of Rorschach scoring categories with preschool children. *Child Developm.*, 1944, *15*, 207-216.

25. Tomkins, S. S. *The Thematic Apperception Test: the theory and technique of interpretation*. New York: Grune & Stratton, 1947.

26. Tomkins, S. S., and Miner, J. B. *The Picture Arrangement Test*: New York: Springer, 1957.

27. Wagner, M. E., and Schubert, H. J. P. *D. A. P. quality scale for late adolescents and young adults*. Kenmore, New York: Delaware Letter Shop, 1955.

28. Weiner, L., Brown, E., and Kaplan, B. A comparison of the ability of normal and brain injured subjects to produce additional responses on a second administration of the Rorschach Test. *J. clin. Psychol.*, 1956, *12*, 89-91.

29. Willis, V., Jr. The relationship of conceptual flexibility to personal adjustment. Unpublished master's thesis, Univer. of Colorado, 1958.

Received December 8, 1958.

Structured and Projective Tests: Some Common Problems in Validation

PAUL E. MEEHL
University of Minnesota

Some of us at Minnesota have been trying to re-think, as objectively as human frailty permits, the principles involved in the clinical use of that structured test with which we are identified.

In preparing for this symposium I discovered that most of the considerations which arise in validating such devices as the MMPI apply with very little modification to those open-ended tasks which have come to be known as the projective techniques. Faced with a choice between rigorously developing a point or two, and offering a more extensive list of suggestions in a brief and dogmatic form with only a sketch of justification underlying each, I have preferred the latter.

In terms of the recommendations of the APA committee on test standards, the great majority of validation problems posed by both structured and projective tests are problems of construct validity. Careful scrutiny of apparent cases of content or concurrent validity will usually reveal that these rubrics are only superficially appropriate. The most obvious example of this is in the use of tests as an aid in formal diagnosis, where there is no point in giving a test in order to predict the verbal behavior of the psychiatric staff. Think of the so-called "hits" and "misses" in a validation study employing routine psychiatric diagnosis as criterion. Almost the sole value of such investigations is their relevance in deciding whether the test can be used as a probabilistic indicator, along with psychiatric judgment or any other "criterion," of the presence of that inner state which is theoretically definitive of a particular taxonomic entity. An important minority of exceptions are those of truly predictive validity, (e.g. prognosis) although even here I would agree with Loevinger's recent emphasis that if predictive validity is to be generalizable over different clinical populations, good construct validity is necessary (6).

It follows that there is very little point in carrying out validation studies unless the non-test indicators are superior, both qualitatively and quantitatively, to the kinds of judgment, rating, or classification which are routinely available. If we want the Rorschach or the Bender or the MMPI to play a role analogous to what the EKG or gold-colloid tests play in organic medicine, we may have to adopt a causal and statistical model more like that of organic medicine. We have nothing strictly analogous to the pathologist's report (a face-valid "criterion"); therefore validation studies must employ multiple assessments of the same construct, numerous internal checks (sufficiently patterned but preserved from contamination), and some procedure for further checks on whether an apparent test miss is actually a miss or only seems to be so. For example: We study the *subset* of cases in which the Rorschach strongly indicated a schizophrenic structure but which were, by the usual concurrent validity criterion, not so diagnosed; then we might determine whether Q-sorts by the residents on each case tended to show a higher Q-correlation with an idealized Q-sort description of "pseudoneurotic schizophrenia" than is observed in randomly chosen non-schizophrenic cases.

Another approach with which we are currently working in our Multiphasic studies utilizes "temporal convergence" to the psychometrics. That

is, recognizing that neither the test nor the non-test indicators (I prefer the term *indicator* to the term "criterion" in such contexts) is infallible, one can at least study the extent to which the non-test indicators and judgments tend to converge to the characterization yielded by the test as more non-test information is accumulated. Remembering the attenuation formula, the demonstration that a set of relatively unreliable clinical indicators converge over time (as a function of repeated and qualitatively diverse exposure to the patient) toward the perception of the patient provided by the test is encouraging, even when the absolute magnitude of the asymptote of this convergence function is mediocre.

A perennial problem in the validation of projective devices (which they share with structured devices when the latter are multi-variate) is the exact role that should be given in validation studies to the so-called "global" approach. While the method of correct matchings has been urged as the solution of this problem, and while I would readily agree that it has a definite place (especially in the early stages), I do not believe it can be defended as the method of choice. It does not present a very true-to-life validation context, since we are practically never concerned clinically to identify "who somebody is." The method does not reveal which variables are doing the work; a device which purports to tell us several different kinds of things may appear better or worse than it actually is on the basis of a small number of crude and clinically uninteresting dimensions to which matchers have learned to attend selectively. Our global thinking must not mislead us into forgetting that in the training of students we *do* teach them to pay attention to certain scores to which the test gives rise and to attach differential interpretive significance to each. That this is done contextually i.e. with reference to the pattern of the remaining scores, does not suffice to justify the matching method, although it does demonstrate the inappropriateness of confining a statistical analysis to the calculation of Pearson correlations between single-test dimensions and single-criterion variables. My own opinion is that Q-sort provides both quantifiability and globality, and is therefore preferable both to the traditional single-variable correlation on the one hand and the method of correct matchings on the other. There are unsolved problems with Q-technique, but our MMPI research has pretty well convinced us that for many clinical validation purposes it is the best method available. I think it unfortunate that the method detailed by Cronbach in 1948 has not been more widely used, although the number of judges required has no doubt militated against its application (1).

It would be nice to have some statistical gimmick for measuring the similarity of two complex characterizations of one person's dynamics and structure. This is not achieved by the method of matching, by single-trait correlations, by Q-correlations, nor, so far as I'm aware, by any method yet proposed. If I describe a person by the usual mixture of genotypic and phenotypic statements, the dynamic point of view sees each facet of the phenotype as an indicator of the genotype; and it is in some sense a more serious mistake to utter the single proposition, "This patient has rigid reaction formations against his unusually strong unconscious hostile impulses" (if that statement is in fact false), than it is to include erroneously a frequently present indicator, e.g., "The patient will be docile in accepting interpretations [because he will perceive disagreement as an aggressive act]" If the characterization were presented in numerical form as estimates of the patient's coordinates in the variables of the indicator or surface trait space, and similar estimates in the underlying causal or factor space, one of the available distance measures might be appropriate (2). Short of that idealized situation, however, one has the feeling that any

numerical measure of the "closeness of fit" between two structural-dynamic characterizations would be excessively arbitrary.

I cannot refrain from a critical comment here with regard to the globality problem. Some clinicians are still confused between the complex psychological processes of the test-interpreter in arriving at his final integrated characterization of the patient, and the statistical problem of validating that interpretation. *After* he has weighed the various possibilities and taken account of the higher order interactions as best he can, the clinician *does* produce a set of utterances which are, taken singly, accurate in varying degrees. It is quite appropriate to determine the number or average amount of errors in the final description by considering the statements individually. Such a procedure, which studies the final result of the clinician's cerebrations, seems to me to be an unavoidable part of validation. This methodological point was made crystal clear by Kelly and Fiske (4) in their exposition of the Michigan assessment program's design, but in spite of the lack of any rational reply to their arguments the principle has even yet not completely sunk in.

A tough practical problem in the validation of multivariate instruments is that of utilizing higher order interaction effects among the test variables. The qualitative merits of the independent construct variable estimates should be excellent, as I've pointed out above, otherwise we are wasting our time. This usually means that the patients have been subjected to intense scrutiny by judges of a high level of clinical competence. On the other hand, the capitalization of random sampling errors increases horrendously as patterning effects are taken into consideration, so that the determination of a function and the assignment of optimal weights requires a large sample. In other branches of applied psychology or even in clinical psychology where predictive validity is really the problem, it often suffices to know a very little bit about a large number of persons; for other purposes, as in carrying out individual psycho-therapy, it suffices to know a very great deal at a deep level about a single person. In validating the configural indicators of a multivariate test, we find ourselves in the unhappy position of requiring to know a *great* deal about a *large* number of persons.

I come now to some considerations relating validation studies to clinical practice. Since the interpretation of multivariate devices involves the test-clinician combination (i.e., mechanical or cookbook procedures do not exist), at what level of clinical skill should the clinicians in validation studies be chosen? One can put several questions, each of which makes theoretical and practical sense. We cannot dismiss negative validation studies on the grounds that the Rorschachs used were not being read by Klopfer but by "ordinary" clinicians of only moderate skill. One empirical property of a clinical device is how much valid information is extracted from it by persons who have achieved that degree of proficiency typical of the population of users. Suppose that only 5% of people who are ABEPP diplomates, who have attended Rorschach workshops and have administered 300 or more Rorschachs under supervision, can arrive at clinically useful conclusions with the instrument; the taxpayer is being somewhat defrauded if the other 95% draw part of their salary on this basis. A second question is, "How much valid and clinically useful information can be extracted by the *most* skillful interpreters of an instrument?" Finally, one may ask, *"How much information is intrinsically contained in the test behavior?"* We are beginning to collect some evidence with regard to the MMPI that *nobody* gets all of the information out (3, 7, 8) and I would assume that to be equally true of less structured devices.

"Clinically useful information" is actually a rather complicated notion.

The *weakest* claim that can be made for a test is that it tells us things about the patient which are true. It is a remarkable social phenomenon that, within an allegedly scientific profession, even this claim has not as yet been clearly supported with respect to the majority of the inferential uses to which our favorite devices are being put. Secondly, one may ask, "How much valid information does the instrument yield which is not readily available from other sources present in any clinical context and who will routinely be making these judgments regardless of whether we use the test?" A third question is, "How much sooner in the time sequence does such incremental information become available to us through the use of the test?" If my test tells me something about the patient which I do not get from the initial diagnostic interview but which I will be able to assess with a relatively high degree of probability by the end of the fifth therapeutic hour, the use of the test is hard to justify.

Finally—the really important question—"What is the incremental efficiency in clinical handling (that is, both in gross administrative decision-making and in the moment-to-moment decisions made in the course of psychotherapy) achieved through incremental information provided in advance by the test?" The difficulties of designing experiments to answer this fourth question are considerable, but it seems to me obvious that the ultimate justification for the clinical use of devices, projective or otherwise, must lie here. I am myself inclined to be somewhat skeptical in this regard, and it is a source of continual amazement to me the extent to which many clinicians take the answer for granted.

In concluding, I cannot miss the opportunity to attack an unfortunate tendency, springing no doubt from over-identification with our favorite devices (combined with a lurking fear that they may not be quite as valid as "everybody knows" they are), to reject any experimental or statistical design which is offered, on the ground that it is artificial or not in accord with actual clinical practice. True, some unimaginative designs have been employed to produce negative results when a more sophisticated approach would have elicited the validities actually present. But I think it is time for some blunt speaking within our profession about the disparity which exists between even the best available studies of validity and the amount of faith—and I use here a theological word deliberately—invested by a large number of practitioners in current methods. Studies such as that of Kostlan (5) in which the incremental information provided by each of several instruments is estimated by showing the effect of *subtracting* that source from the total pattern of data available, are sufficiently close to "real life" clinical practice that no one is ethically and scientifically entitled to ignore the findings. There is far too great a tendency to invoke "clinical experience" as though it were a satisfactory answer to the validation question. Such a mistake could only be made by someone ignorant of the history of science (particularly of the medical sciences) or remarkably naive about those errors in observation, memory, and interpretation to which the human intellect is subject.

REFERENCES

1. Cronbach, L. J. "A validation design for qualitative studies of personality." *J. consult. Psychol.*, 1948, *12*, 365-374.

2. Cronbach, L. J. and Gleser, Goldine C. "Assessing similarity between profiles." *Psychol. Bull.*, 1953, *50*, 456-473.

3. Halbower, C. C. "A comparison of actuarial versus clinical prediction to classes discriminated by MMPI." Unpublished Ph.D. Thesis, University of Minnesota, 1955.

4. Kelly, E. L. and Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor, Mich.: University of Michigan Press, 1951.

5. Kostlan, A. "A method for the empirical study of psychodiagnosis." *J. consult. Psychol.*, 1954, *18*, 83-88.

6. Loevinger, Jane. "Objective tests as instruments of psychological theory." *Psychol. Reports, Monogr. Supplement* 9, 1957, *3*, 635-694.

7. Meehl, P. E. "A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles." *J. counsel. Psychol.*, 1959, *6*, 102-109.

8. Sines, L. K. "An experimental investigation of the relative contribution to clinical diagnosis and personality description of various kinds of pertinent data." Unpublished Ph.D. Thesis, University of Minnesota, 1957.

Received December 8, 1958.

A Study in Validation of the Thematic Apperception Test[1]

WILLIAM E. HENRY
University of Chicago
JANE FARLEY
Parmly Foundation for Auditory Research, Illinois Institute of Technology

The study reported is based upon an analysis of variance design which evaluates the degree of correctness with which TAT interpretations can be identified with the correct TAT subjects, in randomized matching. "Blind" interpretations of TAT's administered to a group of 36 14-15 year old adolescents are systematically compared with similarly "blindly" derived summaries of selected criterion data (objective test scores, summaries of observations and interviews, and Rorschach interpretations). All instruments represented in the criterion data had been summarized independently, but each summary was patterned on the same conceptual framework, agreed upon in advance by each of the clinicians interpreting the various instruments.

The present study, using the data described, follows a model proposed by Cronbach (1). The 36 adolescent subjects were divided into 12 "triads." Criterion material for each triad was separated into statements, and grouped into four general concept areas (Peer Relations, Family, Mental Functioning, and Emotional Adjustment). Expert judges were then asked to match individual TAT interpretations against a series of statement sheets in which criterion statements from three subjects were randomly represented, giving a "t", a "0", or a "?" rating to each comparison. Modifications of the Cronbach design permit the derivation of statements assessing the influence of various effects of the judging procedure, to include the effect of the individual case being judged, the order of the presentation of cases, the effect of the particular criterion instrument being used, the sequence of cases judged, the behavior area being judged, and others.

Since the judging operation comprises two tasks — correct selection of matching items, and correct rejection of nonmatching items — the correctness "scores" of judges were similarly divided. Judges' scores were treated separately, as reflecting on the ability of the TAT to identify descriptions of a particular individual, or on the sensitivity of the TAT in discriminating between individuals. The analysis of variance was undertaken on both identification matchings (I-scores) and upon discrimination matchings (D-scores).

In all instances, hypotheses related to the validity of the original blind TAT interpretations are demonstrably true at a level of confidence very considerably beyond the .01 level. This level of confidence is based on the F-distribution, and makes use of a special index developed by Dr. David

[1] This study was made possible by a grant to William E. Henry from the National Institute of Mental Health, and by a second grant from the Social Science Research Committee of the University of Chicago. It is based upon a dissertation submitted by Jane Farley in partial fulfillment of the requirements for the doctorate at the University of Chicago. The authors wish to acknowledge their great indebtedness to the psychologists who undertook the arduous judging tasks, either in a pilot study or in the main study here reported: Dr. Mary Grier, Dr. Natalie Haimowitz, Dr. Rosalind Cartwright, Mrs. Gisele Mendell, Dr. Hedda Bolgar, Mrs. Harriet B. Moore, Dr. Johanna Tabin, Dr. Sidney J. Levy, Dr. Arthur Hartman, Dr. Robert McFarland, Dr. Irving Leiden; and to Mr. Eric Klinger and Mr. Robert Hendrickson, who assisted at various stages in the preparation of materials for judging and in the statistical analysis; to Mr. Melvin R. Novick, Association of American Medical Colleges, for his statistical advice; and to Dr. David Wallace of the Department of Statistics of the University of Chicago, who served as statistical advisor throughout the project.

Wallace of the Department of Statistics, University of Chicago. The Wallace Index is used in computing each correctness score which is entered in the F-equation, in such a way that the resulting F-value is the percentage by which the correctness of that score exceeds chance. The high level of success thus obtained holds for both the I-scores and the D-scores. Certain differentiations between these scores will be discussed below.

Population

The 36 cases involved were adolescents of approximately 14 years of age, all from the same school in a midwestern community of under 10,000 population. All were "normal," in that they were going to school, living in their own homes. Half of the subjects were girls, half boys.

TAT Interpretations

The basic material to be judged consisted of individual TAT interpretations made by one person upon the records of these 36 adolescents. Interpretations were made blind for each subject, with the exception of knowledge of sex and age.

Criterion Summaries

The basic data against which the TAT interpretations were judged for fit were derived from instruments representing three general types, or levels, of material:

1. Objective Behavior—Sociometric and Guess-Who Tests, Family Relations Questionnaire, Stanford-Binet, academic achievement tests, school grades, and the Emotional Response Test of Bavelas.

2. Subjective Behavior — derived from observations and interviews with teachers, parents, peers in home and in school.

3. Projective Behavior — the Rorschach.

Conceptual Framework

The framework of concepts used was identical for each criterion instrument, and for the TAT (2). It included two gross areas of personality, *Social Interaction*, and *Characteristics of Self*. The Social Interaction elements include peer relations and family relations. The Characteristics of Self elements include mental functioning, emotional reactivity, inner adjustment, the defense mechanisms, and so on. In all major areas, with their various subdivisions, the personality analysis of each instrument was divided still further into 27 statement areas. In the final materials for judging, each statement area was directly comparable. Thus statement No. 1 from any one of the criterion instruments always deals with the "Degree of participation in Peer Groups" (in the Social Interaction area), statement No. 16 was always material related to "Basic Emotional Attitude," (from the Self area), etc.

The 27 statement areas were grouped on the statement sheets according to four larger categories—I. Peer Relations, II. Family Dynamics, III. Mental Functioning, IV. Emotional Adjustment.

Assignment of Cases to Triads

The 18 male and 18 female cases were then randomly assigned to triads, and the position of each subject within the triads similarly randomly assigned.

The Judges and the Judging Procedure

Nine psychologists, all experienced clinicians and projective-analysts, were chosen as judges. For these nine judges, three different presentation patterns were designed, with three judges (a judge "trio") receiving each of the patterns. In this way, replication within trios makes possible a check on the reliability of judging, and each TAT summary is matched with each criterion equally often. The presentation orders were balanced, to avoid or to detect practice effects, or any interactional effects from the sequence of criterion type, sequence of individual

subjects in a given triad, and so on. Scoring instructions for the judges were standardized, and each judge's ratings were made on separate scoring forms.

ANALYSIS OF DATA

We have undertaken two largely independent analyses of variance aimed at the analysis of the judgments of Errors of Identification (I-scores) and at the Errors of Discrimination (D-scores). Within each, we have analyzed the contribution made by some of the various effects of the judging process.

RESULTS

Since the details of the statistical analysis will be reported in a subsequent monograph, we will limit ourselves to a mere statement of our conclusions.

1. *The Validity of the TAT.* We selected, as the evidence for a general validity hypothesis, I-scores which showed a greater than chance relationship between TAT statements and criteria, at a .01 level of confidence. All of the I-score blocks show significant agreement considerably beyond this .01 level. There was no significant variation between the types of criterion instrument (objective tests, interviews, Rorschach) in the agreement of each with the TAT predictions. Although various effects of behavior area, presentation order and sequence, individual subjects, and their various interactions, were for the most part highly significant in their influence upon the scores, even the lowest agreement scores were far above chance in every case. None of the effects or interactions tested depressed the scores of any block below the .01 significance level. Similarly, despite a rather surprising lack of conformity among the judgements of these expert judges, all judges' scores matched the correct "I" ratings at or above the .01 level.

2. *The Discrimination Power of the TAT.* We find an hypothesis about the usefulness of the TAT in expressing the uniqueness of individual personality unequivocally supported by the high Discrimination ("D") scores obtained. These scores, like the Identification scores, are significant considerably beyond the .01 level.

3. *The Scope of the Concepts Relevant for the TAT.* There is no area of concepts, no criterion instrument, no presentation effect which reduces the significance level of the correctness of TAT-criterion matchings. Thus we can claim for the TAT, within the limits of the four broad and 27 detailed concept areas, and within the population dealt with, the wide scope implied by the generality of our hypotheses. However, some differentiations within the various effects of judging were found. By far the greatest single effect on the judging process relates to differences in the behavior area in which the prediction was made (Peer Group, Family Dynamics, Mental Functioning, Emotional Adjustment). This effect is entirely in keeping with expectation. The reason commonly given for assessing personality by a battery of tests, rather than by a single instrument, is that the tests will be focussed upon different areas of personality function. Scores in the four areas mentioned divide themselves neatly into two groups—a Peer Group-Emotional Adjustment category, and a Family Dynamics-Mental Functioning category — with the two sections in each group showing virtually the same percentage of agreement above chance, and the former group noticeably more accurately judged than the latter. It should be noted that the lower efficiency of judgment in these latter areas is only a relative matter, since agreement was significantly beyond chance for all areas. There are several possible explanations of this differential agreement. It may reflect a true differential ability of the TAT. It may reflect the particular preoccupations of the life stage of our population. It may reflect the particular foci of interest of the TAT interpreter.

4. *The "Fatigue" Effect.* Between the first, second, and third sections of the total matchings undertaken by

each judge, there is a consistent, though not extraordinarily large, decline in performance. This applies to both I and D scores, and to all presentation orders, and seems to represent a simple function of fatigue.

5. *Conclusions about the Clinical Use of the TAT.*

a. *There is no single correct way of employing the TAT interpretations.* There was little individual-item-agreement between judges, but each judge made enough "correct" judgments to yield the highly significant agreement figures reported. Judges may arrive at essentially the same interpretive implications by quite different routes; or judges may differ individually in their ability to utilize TAT predictions in different concept areas, or for different subjects. But whatever the route used, all judges arrived at a substantially accurate understanding of each subject, despite great inter-judge variability.

b. *Judges appear to make more efficient use of the projective (TAT) report when the decisions they must make are definitive and detailed,* than when they are relatively simple; i.e., because there is overlap in the behavioral repertoires of individuals, it is easy to make the mistake of including in a subject's potential responses those which are more characteristic of someone else. It is more difficult (and more inaccurate) to exclude from a subject's potential responses those he in fact displays. Conversely, it is less difficult to make the second type of judgment correctly than the first. Although there were more mistakes of the "D" type, as expected, the judges also varied much more widely in their ability to make the "I" type of judgment accurately. We have concluded that the "set" necessary for making fine distinctions (concentration, deliberation, and impersonal detached judgment) are more conducive to evenly accurate diagnosis than the more spontaneous, personally colored approach used in making the less tedious comparisons.

c. *The judges' perceptions were sharpened by consideration of a variety of personality data, but there is a point of diminishing returns.* Although a certain amount of facilitation in judgment is produced by examining the subject from several angles, there is also a "fatigue" effect produced by a superfluity of information, and this effect becomes dominant when the clinician considers many subjects.

d. *Clinicians have their particular 'specialties.'* The accuracy of matching differed with each judge, from individual to individual. There was, at the same time, no clear agreement between judges as to which individuals were judged more accurately. The exact cause of the divergence is not clear, but the crucial variable is apparently strictly personal. That is, it is explained by a difference in the way a judge perceives certain subjects or groups of subjects. It may be that the judge is especially expert at recognizing all hysterics; another judge matches boys better than girls; a third simply finds certain individuals more interesting than others, and is thus more involved with those individuals. In any case, each judge fluctuated in his ability to judge different individuals, and no two judges fluctuated in exactly the same direction.

e. *Instrument.* One negative finding is fully as important as the positive results of any other of the tested variables; this is the nonsignificance of the differences between criterion instruments, and of the instrument-area interaction. It was confidently hoped that the results of this validation study would provide some clue to the place of the TAT in a clinical test battery for personality assessment. Since three different types of instrument were included among the criteria against which the TAT was matched, there was reason to believe that the TAT would show itself to be closer to one than another of these techniques. In

stating the evidence acceptable for confirmation of the validity hypothesis, a "standard" was selected from the three criteria for each area of prediction. This selection was made in the express anticipation of notable area-instrument interaction. The nonsignificance of these effects presents three possible interpretations:

(1) There is actually no difference in level or in accuracy between the reports made from the objective, subjective, and projective tests, as applied to the four behavioral areas.

(2) The TAT is so flexible, and so comprehensive that it taps all levels of all areas equally well.

(3) The interpreters of each of the criterion instruments knew their tests so well that none of them ventured beyond the information legitimately provided by those instruments; consequently, the criterion information should vary in richness and clarity as the information sought was nearer or further from the focus of the instrument; but the statements would not be based on an overextension of the instrument, and would be less likely to be false than a forced statement might be. In this way, judges comparing these statements with the TAT might be required to make decisions on the basis of superficial or meager information, or might be faced with a "no information" item, but would probably not be asked to match the TAT with an incorrect statement. Consequently — assuming a consistent level of validity for TAT interpretation — there might be no significant change in error score (false plus or false zero), but a great deal more uncertainty (question mark).

The first alternative (equivalence of tests) we can discard, since it is contradicted by all of clinical tradition and experience. The second alternative (flexibility and comprehensiveness of the TAT) is very tempting and may well be partially true. However, the "one perfect" clinical test idea does not agree with personality theory, with test theory, or with the experience-based opinions of testers.

The third alternative (limitation of the scope of interpretation in criterion data) —or at best a combination of the second and third — is probably the truest explanation of the unexpected negative finding. We have one ready-made index to the plausibility of this explanation: the hypothesized "meagerness" or "superficiality" of data are matters of opinion, which it would be difficult to verify; but the "no information" items are perfectly definite and verifiable. The distribution of no-information items lends credence to our answer. Using the same criterion for each area, we find the "no information" items distributed according to this pattern. The instrument-area combinations chosen as ideal have the smallest number of N.I. entries. The smallest proportion of N.I.'s given for the observation and interview criterion was in the Family Relations area; the smallest proportion of N.I. items in the Emotional Adjustment areas is from the projective (Rorschach) criterion; the smallest proportion of objective "no information" items is found in the Peer Area. The single contradictory entry in the table is the large number of N.I. items found for the objective criterion in the Mental Functioning area. However, 70 per cent of these data failures occurred on items 14 and 15, which concern creativity and fantasy, and about which the objective data might be expected to have less to say.

REFERENCES

1. Cronbach, Lee J. A validation design for qualitative studies of personality. *J. Consult. Psychol.*, 1948, *12*, 365-374.

2. Henry, W. E. *The analysis of fantasy.* New York: Wiley, 1956.

Received December 8, 1958

What is a Criterion?[1]

EVELYN HOOKER
University of California at Los Angeles

I take the liberty of departing from the general pattern followed by most of my colleagues in this symposium, that is, concern about aspects of the validation process in general as applied to projective techniques. Instead, I shall try to raise some issues about validation within the context of a specific clinical area, namely, homosexuality. This is done with the hope that the issues may be clearer if seen with concrete application; moreover, the theoretical and practical importance of validating projective test instruments for the diagnosis of homosexuality needs no special documentation.

I agree with Professor Meehl that the majority of validation problems are those of construct validity and that careful scrutiny of apparent cases of concurrent validity will reveal that these rubrics are only superficially appropriate. I would extend the case still further, to postulate that studies which restrict themselves to concurrent validity are relatively sterile. Unless our validity studies make clear the nature of the psychological phenomenon being investigated, without depending solely on a diagnostic label of the behavior which constitutes the criterion, we are simply technicians — and not very good ones, at that.

Let me try to make this clear by turning to validation studies of overt homosexuality. The tests which have been most frequently used for the diagnosis of homosexuality are the Rorschach and the TAT. The current validity status of the two instruments can be fairly quickly summarized. In only a bare handful of studies has an attempt been made to validate both of these instruments by comparing overt homosexuals with non-homosexuals, using test indicators called "homosexual indices." Davids, Joelson, and McArthur (1), using the Wheeler (9) homosexual-content indices in the Rorschach, found that these would differentiate a group of overt homosexuals who were college students from a group of non-homosexual "normal" students and also from a non-homosexual neurotic group of students but that they would not reliably distinguish individuals within these groups. Ferracuti and Rizzo (2), in a study conducted in Italy, found that the Wheeler indices, taken as a whole, would not differentiate a group of homosexual prisoners from non-homosexual prisoners but that the presence of any one sign or any two signs would distinguish the two groups. Nitsche, Robinson, and Parsons (7), using a list of content categories similar to Wheeler's, found that the homosexual signs did not differentiate a group of "recently convicted" homosexuals from patients with no known homosexual activity. In my own work (4), the total Wheeler content-indices discriminated between a group of overt homosexuals and a group of non-homosexuals only when the differences in matched pairs were taken into account. Thus, the differences are so small that for clinical purposes they are unimportant. Add to the discrepancies in the findings just cited the fact that, in the groups of records with which I worked, expert Rorschach workers were unable to distinguish homosexual from heterosexual records, even when presented with matched protocols. In none of these studies has it been found that individual records could be validly diagnosed: the criterion "overt homosexuality" obviously covered a "multitude of sins." Kinsey and his associates (5)

[1] This investigation was supported by Research Grant M-839, from The National Institute of Mental Health of the National Institutes of Health, Public Health Service.

have demonstrated the fallacy of assuming that a homosexual is an individual who has engaged in a homosexual act. Unless the extent of heterosexual activity and/or homosexual activity, together with the pattern of desires, is made clear, the criterion becomes ambiguous.

Let us take a look at the results with the TAT. As far as I know, only three studies have been reported in the literature. Davids, Joelson, and McArthur (1), in the study previously reported, used the TAT in addition to the Rorschach. Using TAT homosexual indices gathered from the literature, they found that these, taken as a group, would differentiate the homosexual from the non-homosexual group and that the Rorschach and TAT signs were significantly correlated. Ferracuti and Rizzo (2), using a somewhat similar set of TAT signs, found that they did not distinguish the two groups. Lindzey, Tejessy, and Zamansky (6), also using college students, have found that the TAT indices will differentiate a homosexual from a non-homosexual group but that an uninstructed and unhampered clinician could do a much better job of differentiating than would the objective rater. I found that the TAT and MAPS protocols of the homosexuals previously described in connection with the Rorschach study (4) were so full of manifest homosexual content that neither the judgment of a clinician nor an objective rating of indices was necessary in order to distinguish the homosexual from the heterosexual records. The TAT and MAPS thus would appear to be more valid instruments for detecting homosexuality than is the Rorschach. But let us not be hasty! Lindzey, Tejessy, and Zamansky state (6, p. 73), "It is evident that the variables that function best are precisely those which one would not expect to encounter when dealing with strongly repressed homosexuality or with an overt homosexual who was attempting to conceal this aspect of his behavior." I would like to underscore that remark heavily, not only for the TAT and MAPS but, also, with somewhat less confidence, for the Rorschach. A subject who had given a Rorschach record which, up to this point, gave no indication of homosexuality, produced an oral-genital response which he prefaced with, "I hate to say this, because it will give me away." He had already "given himself away," since he had told me he was homosexual although he was married. He also had told me that he would not have come to me were I living in his home town. He successfully conceals his homosexuality, and is a highly respected professional man in his community. Another said, "I have heard that homosexuals are supposed to see sex in this, but I honestly can't." Test subjects these days are obviously no longer as naive as we like to believe.

The situation is not very encouraging for the clinician who wishes to use any of these three projective instruments for the diagnosis of homosexuality. All of the investigators point out that the use of the signs for individual diagnosis is unwarranted — unless, in the TAT, the subject virtually tells us that he is a homosexual. Why not let him tell us directly? If all one is looking for in the projective protocol is the answer to the question, "Is he homosexual," one may be using an inefficient means of obtaining the answer. If the subject is set not to give the show away, a projective test may not help us much.

Can the situation be improved? Yes, I think it can be, provided we begin to think seriously about the fact that studies of concurrent validity are, in reality, either good or poor studies of construct validity. Or, to put it in a slightly different way, we cannot hope to use projective techniques to answer the question accurately, "*Who* is a homosexual," until we know much more about who or what a homosexual *is*.

In the studies cited, two personality pictures of "the" homosexual have been used. To quote Wheeler (9, p. 123), "If it is possible to accept the ra-

tionale for each item in terms of the aforementioned results, a combined picture of the male 'homosexual' would be: A somewhat paranoid individual with derogatory attitudes toward people, especially women, which is accompanied by a feminine identification. There are indications of anal interests and interest in physical relationships between like beings. There is apparently some preoccupation with sex in general and some autoerotic concern." To quote Lindzey (6, p. 74), "Our tentative findings suggest that the TAT protocols of the individual characterized by strong homosexual tendencies display consistently negative attitudes toward women, a lack of full, rich, and satisfying relations with members of the opposite sex, and occasional reference to manifest homosexuality." A basic assumption which they have in common is that "the" homosexual exists and that he is characterized by attitudes consistent with hatred of, or at least negative attitudes toward, women—a relatively simple construct. I venture to suggest that not only is it too simple but it is not altogether true. Let me hasten to add that many homosexuals can be characterized in this way; but even for them it is a caricature of the complexity of the individuals involved. At the risk of stating a self-evident fact, it is apparent to anyone with knowledge of the wide individual variations in homosexual patterns and life styles, as well as to anyone who has looked at projective test protocols or life histories, that homosexuality is not an entity but is, rather, a multi-manifested phenomenon.

It is precisely here that projective techniques can help us. We need to get beyond the fact that the individual is homosexual, to the kind of homosexual that he is. We are badly in need of fruitful theory which will take into account the multiplicity of variables which appear to be operating. Projective techniques can be valuable sources of constructs for such theory-building. It is not an easy task; and it is one which, I think, will not be accomplished by the development of "homosexual signs." It seems unlikely that signs will help us to diagnose or understand individuals who are as varied as—to quote judges who have interpreted the projective protocols—"an impoverished, restricted, promiscuous, compulsive homosexual, devoid of the capacity for deep feeling"; "a straight heterosexual—except in the bedroom"; "sensitive, independent, autonomous, with strong control of impulses and with a moral attitude toward sex"; "a positive homosexual who goes through life gently, with lots of strength and no hostility, makes strong relationships, and for whom homosexuality is a way of self-actualization"; "a passive-aggressive iconoclast who has libidinized all of his psychological needs in fellatio"; "a paranoid character who has never made up his mind whether he is a boy or a girl, a tormented, sad homosexual who flees from close relations and is at the mercy of his own intense emotions"; "an individual with lots of strength as a man, who plays the assertive role with men, and with tremendous hostility to women"; and, finally, just to bring this arbitrarily to a close, "a homosexual with pervasive feminine orientation, a very passionate person, an aggressive personality with a paranoid base." The most striking finding of the three judges (3) who examined the projective materials was that many of the homosexuals were very well adjusted. In fact the three judges agreed on two-thirds of the group as being average to superior in adjustment. Not only do all homosexuals not have strong feminine identification, nor are they all "somewhat paranoid," but, according to the judges, some may not be characterized by any demonstrable pathology!

It will have become evident by this time that I am not greatly disturbed by the fact that projective techniques are not demonstrably valid means for diagnosing homosexuality. In fact I am rather encouraged by this, because I hope it will force us to re-examine the

much over-simplified picture we have had and encourage us to remind ourselves that the first goal of science is understanding, with prediction and control as secondary to it. It may not be inappropriate to quote here the opening sentence of the APA recommendations for psychological tests and diagnostic techniques (8, p. 1): "Psychological and educational tests are used in arriving at decisions which may have great influence on the ultimate welfare of the persons tested, and of the community." There is probably no clinical area in which psychological tests are used for which that statement has greater import than that of homosexuality.

REFERENCES

1. Davids, A., Joelson, M., and McArthur, C. Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics, and normal males. *J. abnorm. soc. Psychol.*, 1956, *53*, 161-172.

2. Ferracuti, F. and Rizzo, G. B. Analisi del valore discriminativo di alcuni segni di omosessualita rilevabili attraverso tecniche proiettive. *Bolletino di Psicologia e Sociologia applicate*, 1956, *13-16*, 128-134.

3. Hooker, E. The adjustment of the male overt homosexual. *J. proj. Tech.*, 1957, *21*, 18-31.

4. Hooker, E. Male homosexuality in the Rorschach. *J. proj. Tech.*, 1958, *22*, 33-54.

5. Kinsey, A. C., Pomeroy, W. B., and Martin, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.

6. Lindzey, G., Tejessy, C., and Zamansky, H. S. Thematic Apperception Test: An empirical examination of some indices of homosexuality. *J. abnorm. soc. Psychol.*, 1958, *57*, 67-75.

7. Nitsche, C., Robinson, J., and Parsons, E. Homosexuality and the Rorschach. *J. consult. Psychol.*, 1956, *20*, 196.

8. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull. Suppl.*, 1954, *51*, Part 2, 1-38.

9. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *J. proj. Tech.*, 1949, *13*, 97-126.

Received December 8, 1958

Validity and Methodology in Projective Tests

NORMAN L. FARBEROW
V.A. Center, Los Angeles

The reflections in this paper stem from a study by Dr. Shneidman and me on "TAT Heroes of Suicidal and Non-Suicidal Subjects," a report of which appeared in the June 1958 issue of the *Journal of Projective Techniques* (10). This was a study which started out as a clinical investigation of suicidal individuals as inferred from their reactions to the TAT, but which ended up as an examination of the validity and methodology of the techniques used. This is not the occasion to go into details about that experiment, but in order to report these reflections, let me briefly review some aspects of the study.

In essence, ninety-three TATs were gathered into seven groups which were then compared with each other. The categories consisted of three groups of males, normal, neurotic, and psychotic, who were *non*-suicidal; two groups of males, neurotic and psychotic, who had *attempted* suicide; one group of neurotic females who had *attempted* suicide; and one group of neurotic males who had completed suicide. The technique used was one devised by Ira Friedman in which, instead of the usual procedure of rating for the attributes of the story-teller, the attributes of the TAT *heroes* of the TAT protocols were sorted (3, 4). These were rated—completely blind—by means of an 80-item, nine-step, Q-sort procedure by one judge, in this case, Dr. Friedman himself.

It was our hope that the sorting for the attributes of the story-hero rather than for the characteristics of the story-teller would permit us to eliminate some of the harrassing problems of unreliability and to come closer to grasping some of the essential differences in the thematic productions of suicidal and non-suicidal subjects. We do not wish to say that our analysis of the statistical results were completely disappointing, but suffice it to say that on viewing the results one of us went into a sharp reactive depression. In reflecting back to our feelings at the beginning of the study, it seems that we covertly had the naive hope that the results would be something as follows: that our reliabilities would be perfect; that the within-group correlations would be plus one, indicating complete homogeneity within each group; and that the between-groups correlations would be around zero, indicating that these homogeneous groups were, indeed, different from one another. We could then describe in some detail those clinically significant attributes which nature, the TAT, and statistics had shown to characterize each of the seven groups. Clinicians would then have new and practical clues to sharpen our diagnostic acumen and to enable them to distinguish among individuals who have attempted suicide or who will commit suicide. Our IBM equipment, alas, betrayed us. It appeared that none of the seven groups was essentially homogenous and that while our judge was able to make essentially correct over-all judgments better than chance as to the sex of the subjects, and as to normal versus disturbed (but not suicidal) subjects, the differentiation between suicidal and non-suicidal status yielded no better than chance expectation. In thinking through the results we came to the conclusion that the study had some interesting implications for the validity and methodology of picture thematic tests.

We were faced with the inescapable results that the TAT, as it had been employed in our study, could not differentiate between groups of suicidal and non-suicidal subjects. The possibility that there were no essential differences was a conclusion we could not

comfortably accept, having had the benefit of a great deal of clinical and empirical observation behind us from over a decade of intensive research pre-occupation with the subject. Our conclusion was that not only was the nature of the phenomenon we were attempting to measure so complex, but that the nature of the stimuli used to evoke such differences were either so varied, or even inappropriate, that the differences must, indeed, have been masked.

In terms of the phenomenon, or better, the phenomena, of suicide, we felt we were under no misconceptions as to the complexity of the subject we were studying. To paraphrase Dr. Hooker, it is extremely difficult to know who *is* suicidal until we know who the suicidal person *is*. As an example of the complexity, we need only point to such varied and desperate individuals as the young teen-ager who blows off his head on his jilting girlfriend's porch; the older man, widowed and with a terminal disease, who gently slashes his wrists and lied down in bed slowly to bleed to death; or the schizophrenic girl who attends to an hallucinatory voice that the first step down from the sixth story window of the hospital is not so high after all. For still more examples, we may look among the attempted suicides, where we may have the aborted successful attempt in which the individual is saved only by a freakish accident, or the genuine attempts with various degrees of a "gamble with death" (11), or the gestural attempt, where the intent to die is completely absent. It becomes obvious that the motives or the patterns of motives may run the gamut of those previously conceptualized, such as those offered by Durkheim, Freud, Jung, Menninger, Kelly, and others, and includes many, we are sure, yet to be conceptualized, we hope, by us and by others working with this problem. In our search for further constructs in the understanding of suicide, our disappointment served to remind us again that neither we nor our tests are omnipotent, and that we cannot reasonably expect universal and eternally perfect correlations with aspects of nature we have only incompletely conceptualized.

Our questions at this point were old familiar ones. For example, we even debated the basic question of whether it was possible to identify or to predict an overt act or the state of mind which produces such behavior, from fantasy productions aroused in a special set of conditions, conditions which were usually very carefully arranged beforehand to eliminate any behavior other than the production of fantasy. I will not burden us with reports of the experiments on the relationships between test performance and behavior other than to remind us that they have been both confusing and frustrating. Dr. Bert Forer, Editor of the *Journal of Projective Techniques,* expressed essentially our feeling when, in commenting on the papers in last year's Symposium on Research in Projective Techniques, he said, "There is a good deal of psychological substance between drives and behavior." Nevertheless, while we are aware that the expression of a motive in one context or at one level of psychological organization has not been too highly correlated with its manifestations at another psychological level, this basic assumption must be made or indeed there can be no justification for projective tests. In other words, there must be an essential relatedness or equivalence of behaviors; the task then becomes one of finding behaviors with the highest equivalence. This becomes simply another way of stating the basic problem of validity: will the test data yield meaningful information about the behavior being examined.

It becomes apparent from the above that in the use of projective tests the nature of the stimuli employed in the investigation of any set of phenomena assumes proportions of major importance. This attention to the nature of the stimuli would fall within that special aspect of validity suggested by Dr.

Shneidman in his five-part delineation of the problem of validity, specifically, Validity For What.

In reflecting on the nature of the stimuli we had employed in our study, the closely related questions of globality and of relevance seemed important. The evaluations of an entire TAT protocol in terms of composite judgments for the identification or prediction of particular kinds of behavior have generally led to poor results. The TAT, like most projective tests, was conceived of as a global test, yielding many data about many facets of personality, with, as we have painfully learned, much depending upon the skill of the clinician-interpreter. Today, much of the trend seems to be toward devising instruments for measuring the individual constructs which go toward making up the global personality, a trend to which we wish to add our voice. Cronbach and Meehl in their article on Construct Validity in the *Psychological Bulletin* in 1955 (2) stated that global interpretations require validation of the separate constructs implied or inferred in each separate statement about an individual. Bindra in his article entitled "Psychotherapy and Recovery from Neuroses" (1) suggested that globality may be all right for the clinician, but that the researcher must concentrate on "Truth" in small portions, chewable bites, and digestible bits. Kelly, in his review chapter on assessment in the 1955 *Annual Review of Psychology* (6), lists tests for individual constructs, such as anxiety, dependency, and other familiar terms. In addition, the relevance of a test, such as the TAT, or Rorschach, for measuring a phenomenon also rests squarely in the middle of validity, the Validity of What question. The TAT, by the very nature of its construction, taps a wide range of many facets of personality and could not hope to encompass an adequate appraisal of the polarities of each of the psychological constructs invoked. Other tests measuring such polarities within the individual constructs are needed, not to supplant the TAT, but rather to supplement it.

At the beginning of the present study it seemed to us that the technique of evaluating the attributes of the hero offered an unusual opportunity to stick close to the given data and thus to obtain interjudge reliabilities based on phenotypes. What we had not anticipated was that one cannot collate or average or add heroes. Consider: the clinician's usual diagnostic task is comparable to seeing a play, for example *Hamlet,* and then divining Shakespeare's personality attributes. In the hero-oriented approach, the task is to see *Hamlet* and then describe, in some systematic way, Hamlet's personality characteristics. But, if on a single day the clinician were to see *Hamlet, Love's Labor Lost, Othello, Measure for Measure, King Lear,* and *All's Well That Ends Well*—how could he then give a *single* description of the chief protagonist of *all* these plays? Unfortunately, this was the judge's task in the present study.

The recommendation at this point is an obvious one (and one which we intend to pursue); namely that the protocols be separated into their respective stories and judgments be made across the protocols in terms of single stories, that is, all the heroes described in the stories given to Card 6BM would be judged, the heroes of 7BM, 13MF, etc., and then the correlations could be made. Further, we would intend to design specific sets of picture thematic stimulus materials, to investigate hostility feelings, suicidal feelings, etc.

Our suggestion, then, is for a modification of the stimulus data, using the TAT principle. We all know that the TAT has been manipulated in several ways. The first way that occurs to one is that Murray himself devised the TAT materials so that there was contained within them a continuum between highly structured and highly ambiguous stimuli. The second way in which the TAT principle has been manipulated is by variation of the content of the test materials for particular

purposes, i.e. the creation of special sets of TAT like cards for specific kinds of subjects. There are, for example, TAT modifications for Negroes, for Navajoes, for Utes, for paraplegics, for Michigan adolescents, Korean prisoners-of-war, and even for some animals like the CAT. What we want to suggest is a third dimension along which the TAT principle can be varied; that is, for the measurement of specific behavioral and emotional constructs. We believe that what seems to be called for are *specific sets* of projective materials which are devised especially and used experimentally to explore the *particular* continua of personality functions. Specifically, we suggest, that special sets of picture thematic materials be constructed along such continua as hostility, homosexuality, passivity, suicidal potential, etc. This suggestion is in line with previous work by Little and Winder (8), and by Lesser (7). In concentrating on the area of interest to us, we envisage the construction of a set of cards which would fall along the continuum of suicidal-fantasy evoking stimuli, in which is contained one card (among several) that has for practically all subjects no suicidal card-pull, and at the other end another card that has for practically all subjects a prepotent suicidal card-pull. As Murstein has similarly stated in his article "Nonprojective Determinants of Perception on the TAT" (9), it is important to know the non-projective determinants of perception on any stimulus so that the amount of projection may be evaluated. Suicidal fantasy on one card might indicate considerable distortion of the stimulus properties, while on another card reflect only considerable respect for the objective stimulus. Our interest in administering such a series to a subject would be in his limen or threshold—at which card did he begin to talk about suicide—as well as in his total suicidal score. We are exploring the possibility of including within the scale some other measure which might yield an index of the possibility of acting on the suicidal tendencies indicated, or, to give it a psychoanalytic conceptualization, evaluate the adaptive aspects of ego functioning. The specific determinant might be anxiety (construct: add anxiety and one gets agitation, a significantly dangerous suicidal state) or guilt (construct: guilt coupled with fear stimulates expiatory hopes and activity, also a significantly dangerous suicidal state.)

These suggestions are not meant as "practical" in the sense of our rushing to construct another test for the Psychological Corporation catalogue. Dr. Friedman, in his critique of our paper published in the September 1958 issue of the *Journal of Projective Techniques* (5), after indicating that our suggestion appears to be an approach which might reduce the thorny problem of collating, averaging, or adding heroes as well as an attempt to put subjects on the same measuring scale with reference to particular personality variables, then goes on and lists some misgivings about this notion, which essentially have to do with quantification of the data as opposed to the more global approach. Parenthetically, we might say that Dr. Friedman sent us a copy of his Critique before it was published, and we were asked if we wished to answer it. We were hardly in a position to rebut his paper inasmuch as we agreed with (or had stated) everything which he said. However, we see no reason why we should not pursue this avenue of approach, if for no other reason than to see where it might lead. In this sense, we find the current efforts in the field of thematic testing vastly exciting in that they continue to suggest new procedural techniques as well as new conceptualizations of human behavior which may be useful in the exploration of personality.

REFERENCES

1. Bindra, D. Psychotherapy and the recovery from neurosis. *J. abnorm. soc. Psychol.*, 1956, *53*, 251-254.

2. Cronbach, L. J., and P. E. Meehl. Construct validity in psychological tests. *Psychol. Bull.*, 1955, *52*, 281-302.

3. Friedman, Ira. Objectifying the subjec-

tive: a methodological approach to the TAT. *J. Proj. Tech.*, 1957, *21*, 243-247.

4. Friedman, Ira. Characteristics of the TAT heroes of normal, psychoneurotic and paranoid schizophrenic subjects. *J. Proj. Tech.*, 1957, *21*, 372-376.

5. Friedman, Ira. A critique of Shneidman and Farberow's "TAT heroes of suicidal and non-suicidal subjects." *J. Proj. Tech.*, 1958, *22*, 281-283.

6. Kelly, G. A. Theory and technique of assessment. In *Annual review of psychology*. Stanford University Press, 1955.

7. Lesser, G. S. Conflict analysis of fantasy aggression. *J. Pers.*, 1958, *26*, 29-41.

8. Little, K. B., and C. L. Winder. A Scaled Thematic Test for aggression fantasy in children. (In preparation).

9. Murstein, B. I. Nonprojective determinants of perception on the TAT. *J. Consult. Psychol.*, 1958, *22*, 195-198.

10. Shneidman, E. S., and N. L. Farberow. TAT heroes of suicidal and non-suicidal subjects. *J. of Proj. Tech.*, 1958, *22*, 211-228.

11. Weiss, J. M. A. The gamble with death in attempted suicide. *Psychiat.*, 1957, *20*, 17-25.

Received December 8, 1958

Problems in the Validation of Projective Techniques

KENNETH B. LITTLE
National Institute of Mental Health, Bethesda, Maryland

We are discussing today the general issue of the validation of projective techniques. However, I will limit my remarks to the effective validity of such instruments rather than attempt to cover all forms of validation. "Effective validity" may be subsumed under predictive validity but is restricted to the class of useful predictions rather than any and all predictions.

In the clinical situation, where most projective techniques are administered, validity has a most pragmatic air. The construct validity of the instrument may be of theoretical interest to the psychologist and the concurrent validity an issue of some curiosity, but as far as the case in hand is concerned, one usually wishes to make meaningful and useful predictive statements about a patient. When one administers a Rorschach or a TAT to such a person, the implicit or explicit questions are of a rather prosaic sort: "Will this patient attempt suicide?", "Does he have some form of organic impairment?", "Would psychotherapy be an appropriate form of treatment?", and the like. The questions are those the psychologist, as a psycho-diagnostician, was hired to answer.

Other information may be, and usually is, provided in the test report. And certainly there is little expectation that the answers to the questions will reach such a degree of specificity as, "This patient will jump out of the ward bathroom window at 5:54 a.m. on Thursday." The report should include, however, statements about the subject that raise the probability of making correct assumptions as to his future behavior. The predictions involved may be quite indirect and dependent upon fairly complex chains of inferences but if personality assessments are to be other than exercises in composition, some information with implications for behavior must be included. It is the accuracy of the significant predictions, the correctness of the answers to the perennial questions, that may be called effective validity.

After this preamble, it is somewhat embarrassing to have to say that the published evidence on projective techniques indicates that they have either zero or, at best, very low positive effective validity indices. Even in those studies with the most positive of results, correlations are of an order of magnitude which make predictions for the individual largely a waste of time.

This distressing state of affairs has been with us unchanged for a rather long period of time. During the 35 or so years in which, for example, the Rorschach has been subject to systematic investigation, the only noticeable change has been in the sophistication of design and analysis. Negative or ambiguous results, however, are just as disheartening when determined by an F ratio as when determined by a critical ratio.

The situation as described has been pointed out by several people, e.g., Cronbach, (4) so I will not spend time in citing evidence. Instead I would like to consider the questions of "Why?" and "What do we do now?". A definitive answer to either of these would still require far more time than is available in this paper so I will perforce be quite arbitrary about what I present.

An answer to the question of why our projective techniques fail so miserably at practical predictions is obviously required before any rational solutions can be proposed. However, the determinants involved in the failure are numerous and many and quite difficult to eliminate. The fact that be-

havior is a function of both the organism and the situation means that any prediction based upon evaluation of the former alone will always be a contingent one. Since the number of possible alternative stimulus situations is astronomical the problem of raising the validity of our instruments seems an impossible one. Fortunately humans as well as rats are capable of, and facile at, generalization, so it is possible to reduce the number of stimulus situation categories sharply. We are able to say that one man's candy is another man's brandy and call both consummatory responses, or oral behavior, *etc.* This inevitably leads to a loss in precision in the prediction of specific behavior so we find ourselves anticipating alcoholism when obesity in fact occurs. However, correlations of less than unity are a fact of life with which psychologists have already learned to live.

However, there are other issues having to do more directly with the assessment process and with the instruments used. In our limited time, I wish to talk about two of these, one quite speculative and the other rather concrete. The speculative issue has to do with the interpretation of the basic postulate of projective testing, that of psychic determinism. The term "interpretation" is used deliberately since few people now argue that personality factors have no effect whatsoever on responses to projective technique stimuli. The rub occurs when we assume that all or a majority of the responses are determined by dominant and enduring characteristics, equally operative in the every day life of the individual as in the testing situation. The experienced tester allows for set, defensive factors unique to the particular situation, and the like in his interpretations. Then at this point he often assumes that what remains represents the regnant motives, ego-protective mechanisms, behavior style and so forth that characterize the individual in all situations. I wonder if this is necessarily the case. Perhaps all that is revealed is a micro-personality of limited extent.

This amounts to a questioning of the unity of the personality, its Gestalt quality. It does not deny psychic determinism but merely raises the query as to whether that aspect of the psyche involved in responding to semi-structured ambiguous stimuli is of any practical significance in the rest of man's behavior. During great stress when the individual is fighting for physiological or psychological survival, all forces are mobilized toward a common end and common response patterns may be apparent in all or most all situations. Under less extreme conditions perhaps the individual is far more differentiated than we expect, with a wide selection of response roles. Even in physiologic functions such an effect may be observed; systolic and diastolic blood pressure, for example, correlate .60 in subjects suffering from combat fatigue but the relationship drops to .06 after recovery (10). A similar phenomenon may occur among psychic functions.

The above, as stated earlier, is obviously speculative; there are, however, specific aspects of our assessment instruments towards which we can point with some assurance in the attempt to determine why they fail. These have been examined at some length by Anastasi (1), Cronbach (4), MacFarlane and Tuddenham (7), and others, so I will discuss only the sampling problem.

In the Rorschach one has ten stimuli and from ten to sixty or so responses; in the TAT twenty pictures eliciting hopefully twenty stories; in other projective techniques a varying number of stimuli but of a similar order of magnitude. With these stimuli we secure a small sample of the behavior of the individual and assume it is representative of his total population of imaginative responses. Then to make life even more difficult we attempt to "measure" five to seventy-five different variables with this one sample. The wise tester soon learns to restrict his

assessment endeavor but typically we find that we are still assigning values to the variables we *do* use on the basis of one or two responses. Since these few responses are simply a sample of a number of potential responses to the stimuli, the reliability of the measurement is quite questionable. The projective technique interpreter is in a position analgous to that he would be in if one took a random sample of 50 items from the MMPI, permitted the subject the option of answering them in any fashion he chose (including not answering them at all) and then asked the psychologist to describe the MMPI profile. Depending upon the sample of items, the subject's choice of answer categories, and the tester's knowledge of the MMPI scales, there might be an occasional fairly correct reproduction of the whole profile, a slightly more frequent fair approximation of one or two scale scores and, most probably, an overall accuracy index that resembled strongly the validity co-efficients for projective techniques.

Since others on the symposium have discussed this point also, I would like to turn to my second question, *i.e.*, what do we do now. This discussion I will limit to some suggestions regarding increasing the precision of measurement of projective techniques.

There are several different ways in which this might be done—all relatively painful. The first, which continues the idiographic global approach, would be simply to increase the length of the current instruments — a Rorschach of 50 cards, a TAT of 100, for example. It is theoretically possible in this fashion to secure enough replicated information about a subject to determine the interrelations among personality characteristics and to enable us to specify the links between personality structure and behavior.

I do not have much hope for such a procedure; increasing the size of our global instruments will increase proportionately the amount of irrelevant information to be handled, placing a still greater load on the intellectual and experience level of the clinician. Unless we are to increase training time indefinitely it is almost mandatory that our instruments do a certain amount of pre-digesting for us, sacrificing coverage for precision. As Tomkins and Miner describe it (9) we must be prepared to throw away information in order to raise the probability of correctness of that we do obtain.

If we adopt such an austerity program, there appear to be two modes of approach which might be used. The first way might be considered analogous to the item analytic techniques used in objective test construction. The literature on the existing techniques can be searched for stable and hopefully useful determinants of responses to projective stimuli and special scales constructed for their measurement. After construction of the scale, there will be a long and tedious period of establishing the empirical correlates of the scale. A good example of this method is Barron's (2) M scale for measuring the human movement response to ink blots. If such single determinant scales are too molecular, then additional scales can be constructed for configural combinations; the essential point is that there be sufficient stimuli for reliable measurement.

The second approach is the construction of projective instruments designed to measure specific personality variables within certain theoretical frameworks. The Blacky Test (3) incorporates this idea to some extent but still attempts to measure too many variables with too few items. For precision of measurement, a minimum of about five reasonably independent projective "items" seems essential. Some of the sentence completion tests (Forer (5) and Rotter and Rafferty (8)) might be considered illustrations of this approach. McClelland's use of the TAT pictures for assessing Need Achievement also is a method of measurement that can be considered of this type.

A final example of a thematic projective technique devised for measurement of a specific personality charac-

teristic is one developed by C. L. Winder and myself. It is similar in its general construction to an instrument developed quite independently by Lesser (6). This test which has no formal title is designed to tap two aspects of the fantasy life of boys—fantasy aggression and fantasy aggression anxiety. In constructing it we started with 100 pictures drawn by an artist to represent a variety of boy-boy, boy-girl, and boy-adult interactions and with varying degrees of pull for aggressive response. The subjects wrote stories to the pictures under general TAT type instructions. Reliability of scoring is better than .92 and split half reliabilities (alpha coefficients) range between .71 and .88 depending upon the type of aggression fantasy, *e.g.* boy-boy, boy-adult, being assessed. Since this is but the first round of refinement we hope to increase those reliability figures somewhat.

Such instruments as this last demonstrate very well the advantages and disadvantages of the projective inventory approach. A tremendous amount of information is sacrificed in the interest of precision, without, it should be added, with any guarantee that the resulting scores will be any more useful for predicting behavior than the global assessment. But since the global methods do not appear to work, we must, of necessity, try something else. Such inventories, moreover, do not permit the integration of multiple factors into a unique personality description, at least when administered singly. However, since it is my belief that unique descriptions that do not generate accurate predictions are merely works of art and not clinically meaningful, I think the gamble is worth taking.

There are other experimental variations of projective techniques such as the PAT (9) which time does not permit me to discuss. They represent, I think, a very healthy trend toward exploration of all possible methods for personality assessment. If projective techniques are to be other than examples of the ingenuity of psychologists and their interpretation other than a measure of the clinicians verbal ability, then some modification of our present instruments is essential.

References

1. Anastasi, A. *Psychological testing.* New York: Macmillan, 1954.

2. Barron, F. Threshold for the perception of human movement in inkblots. *J. consult. Psychol.*, 1955, *19*, 33-38.

3. Blum, G. S. *The Blacky Pictures: A technique for the exploration of personality dynamics.* New York: Psychological Corp., 1950.

4. Cronbach, L. J. Assessment of individual differences. In Farnsworth, P. R. (Ed.), *Annual review of psychology.* Stanford, Calif.: Annual Reviews, 1956.

5. Forer, B. R. A structured sentence completion test. *J. proj. Tech.*, 1950, *14*, 15-29.

6. Lesser, G. S. Conflict analysis of fantasy aggression. *J. Personal*, 1958, *26*, 29-41.

7. MacFarlane, J. W., and R. D. Tuddenham. Problems in the validation of projective techniques. In H. H. Anderson and G. L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice Hall, 1951.

8. Rotter, J. B. and J. E. Rafferty. *Manual for the Rotter Incomplete Sentences Blank, college form.* New York: Psychological Corp., 1950.

9. Tomkins, S. S. and J. B. Miner. *The Tomkins-Horn Picture Arrangement Test.* New York: Springer, 1957.

10. Wenger, M. A. *Studies of autonomic balance in Army Air Forces personnel. Compara. Psychol. Monogr.*, 1948, *19*, No. 4.

Received December 8, 1958

Further Check on the Diagnostic Validity of the Ames Danger Signals

LOUISE BATES AMES
(With the assistance of MARJEAN KREMER)
Gesell Institute of Child Development

An earlier study of child Rorschach responses (1) has demonstrated that many kinds of response customarily considered as indicative of disturbance or even of pathology in the adult, occur not only benignly but even normatively in the child subject. Particularly at such ages as 3½ and 7 years, the apparently typical Rorschach response includes several conspicuous features which would suggest disturbance in the adult.

Nevertheless, in spite of the apparent normality of many traditional danger signs when such responses appear in the child's record, we know that the clinician who works with children does depend on the Rorschach for clues as to the nature and degree of disturbance in presumably disturbed child subjects. Thus we face the problem — which kinds of responses customarily considered to be danger signals actually do characterize the response of children judged on other measures to be "disturbed" children.

We ourselves tentatively identified a list of thirteen signs which seemed to us potentially indicative that a subject was in real difficulty. The first attempt at validating this list of danger signals is that of Elkins (3). Of the thirteen "Ames danger signals" suggested by us, Elkins dropped four as involving subjective judgments, and then increased the number of signs to be tested to a total of thirteen because of the necessity of treating the parts of four complex signs as separate signs. Her list, then, included the following:

1. Preponderance of tree or leaf responses (Preponderance defined as over 50% of R).
2. F% exceeds 75.
3. More than one pure color response.
5A. More than one response involving differentiated shading.
5B. More than one response involving diffuse, dysphoric shading, i.e. Clob.
6. More than one inanimate movement response.
8. W% less than 40.
9A. Dd% over 15. Dd% equals Dd, S, and Do responses over total R.
9B. More than one X response.
9C. More than one S response.
10A. A% over 60.
10B. No human response (i.e. no H; Hd not counted).
11. R less than 10.

Elkins selected a group of forty presumably normal 8, 9 and 10 year olds matched for sex, age and I.Q. with a group of forty children who had been referred to the Institute for Juvenile Research in Chicago as being in need of child guidance services. The Rorschach was given to each of these subjects.

Following careful procedures, Elkins checked these records for presence or absence of the suggested "danger signs", and then compared the total scores for the clinic and nondisturbed groups, by means of a variation of McNemar's Formula 26 for the evaluation of the difference between two correlated means. The resulting figures showed that "the clinic, or presumably disturbed, group had significantly *fewer* danger or maladjustment signs than did the nondisturbed group."

A more detailed analysis by individual signs indicated that "most of the Ames danger signals are not useful for differentiating between clinic and nondisturbed samples of 8 to 10-year-olds because nearly half of the signs are either characteristically present in a sizeable proportion of both normal and clinic cases, or else are almost never found in the records of either... The most promising signs which discriminate in the expected direction are pure C > 1, more than one diffuse,

dysphoric shading response, and m > 1. These signs, like contamination for schizophrenia, are apparently diagnostic when they occur, but have such limited incidence as to be of little use for routine screening purposes."

Such extremely negative findings suggest to us two possibilities. The first is that even to a greater extent than we had suspected, the kinds of Rorschach response often considered, especially in the adult, to be indicative of disturbance, have a benign significance when they occur in the records of children. (Our earlier study had, in our opinion, ruled out as danger signs many kinds of response often considered to be dangerous, but we had nevertheless offered tentatively as possible danger signals the thirteen signs tested by Elkins.)

An alternative possibility, and one which we favored, was that Elkins' clinic population of forty subjects, whose records did not show a significant number of our suggested danger signs, were perhaps not seriously disturbed children. Elkins herself feels convinced that "the clinic population tested represented a really disturbed group". In fact she is more convinced of this than that her "nondisturbed group was an ideally normal population."

However, it has been our finding in clinical practice that a large majority of children brought to us for clinical help, are not what we consider seriously disturbed children.

The present study duplicates Elkins' investigation, with the possible exception that we took great care that our disturbed population was made up only of subjects whom we had previously classified as clearly disturbed, and that we used boys only.

Subjects and Methods

Records of all clinical cases between the ages of 6 and 12 years, seen at the Gesell Institute of Child Development during the five years from June, 1953 to June, 1958 were reviewed by means of reading over case-conference notes. All cases where the case-conference exclusive of the Rorschach, which was not considered, indicated "emotional instability", "emotional disturbance", "need for therapy", were then checked with the summary of these cases written by Dr. Frances Ilg, Director of the Clinical Service. Only those cases were selected which were labelled by Dr. Ilg as clearly disturbed.

Since in this total group of cases, records of only six girls were included, it was decided to use boy subjects only. In order to make up a total of fifty cases, records from the six month period from January to June, 1953 were included. This final total of fifty subjects whom Dr. Ilg considered to be clearly unstable or disturbed, were thus chosen from a total of approximately three hundred records of boys of ages between 6 and 12 years.

Only children with I.Q.'s rated as average or better were included.

A control group of fifty presumably normal subjects was chosen from among our normal research population as follows. All disturbed cases were checked for age, I.Q., and father's occupation. Then a research assistant (MJK) selected from among our normal population cases which matched the disturbed cases for age, I.Q. and father's occupation.

SES and intelligence ratings for the two groups are given in Table I. Similarity of the two groups is marked. The largest numbers of subjects in both groups are from the professional or semi-professional classes. Only 7 of the disturbed subjects and only 8 of the normal group fall below Class III, that of clerical, retail business or semi-skilled occupations. As to intelligence, largest numbers in both groups fall in the average category, next most in high average. Only 8 of the disturbed subjects, only 6 of the normals are unclassified — thus possibly below average.

During the selection of these two groups, the normal and the disturbed, the Rorschach records were not consulted nor considered. That is, neither the "disturbed" nor the "normal" sub-

TABLE I—Socio-economic Status and Intelligence of Subjects

Class	No. normal subjects	No. disturbed subjects
	Socio-economic status	
I	16	16
II	15	16
III	11	11
IV	—	—
V	5	6
VI	1	—
VII	—	—
?	2	1
	Intelligence ratings	
Very superior	1	2
Superior	9	7
High Average	15	9
Average	19	24
Unclassified	6	8

jects were selected on the basis of their Rorschach performance. Once selected, however, both groups of cases were then checked for the occurrence of possible "danger signals" in the Rorschach records. Our original 1952 list which, in a slightly modified version (3, p. 283) was tested by Elkins, is as follows:

1. A preponderance of leaf or tree responses, particularly if given perseveratively, after 4 or 5 years of age. Also static perseveration of any simple one-word response.
2. After 5 years of age, more than 75%F.
3. More than one pure C is suspect.
4. Color naming is suspect after 5 years.
5. More than one shading or clob in a normal length record.
6. More than one m.
7. Contamination after 5½ and confabulation after 7.
8. Less than 40 to 50%W.
9. More than 10 to 15%Dd. Especially more than one Do or more than one S.
10. Over 60%A. Less than one H.
11. After 3½ years, a total of less than 10R.
12. Position responses.
13. Other behaviors which are at least suspect are the giving of extremely bizarre responses, inability to cope at all with the test, excessive manipulation, or particularly odd manipulation.

Subsequent experience with the Rorschach test in children has led us to make the following modifications of this list:

1. "Preponderance of leaf or tree responses" was modified to read, "Three or more leaf or tree responses, or 15% or more of total content."
2. 75% F or more was changed to the probably more significant F%>F+%.
3. "One or more C" was changed to "Any C or color naming."
4. "More than one shading or Clob in a normal length record" was modified to "Any variation of Clob or C'."
5. "More than one m" was changed to "Three or more m or m > M + FM.
6. Area was omitted as a criterion except for "25% or more Dd or more than one Do."
7. "Over 60%A" was raised to "Over 70%A."
8. "Less than one H" was changed to "No H; or Hd: H is 2:1."

The following items were added:

1. "Elimination, sex or reproduction responses."
2. "Three or more anatomy, or 15% or more of the total content; or F— anatomy."
3. "Personal reference or more than one response suggesting trouble, as old, broken, decayed, damaged objects."

With these modifications, we arrive at the list of possible "danger signs" listed in Table II. This list, because of these changes, is slightly different in detail from the list which Elkins attempted to validate. However, it is similar; and since our aim is to discover what kinds of Rorschach responses are danger signals, and not merely to "defend" our original list, these variations in the original list seem permissible.

Table II lists all proposed danger signals and indicates the number of instances, out of 50 normal and 50 disturbed subjects, paired, in which both of the pair, neither of the pair, only the disturbed subject or only the non-disturbed subject showed the sign. It also indicates what percent of each group show each sign, and includes a statistical measure of significance.

Mean number of these proposed signs in our normal group is 1.6; in our disturbed group, 5.0. There is a statistically significant difference ($t=15.90$) between these figures. However, as Table II shows, some of the individual measures within this list are more effective than are others in distinguishing between normal and disturbed subjects.

TABLE II—Incidence of "Danger Signals" in Matched Pairs of Clinic and Nondisturbed Subjects

Danger signal	Sign occurs in Both of pair	Neither of pair	Only non-disturbed	Only clinic	Chi-square[a]	Percentage of occurrance Non-disturbed	Clinic
1a. Less than 10R	5	31	4	10	1.79	18%	30%
1b. Excessive manipulation	0	38	0	12	*10.08***	—	24%
2a. More than 25%Dd	1	41	3	5	.12	8%	12%
b. More than 1 Do	1	39	2	8	2.50	6%	18%
3. F% exceeds F + %	2	29	2	17	*10.31***	8%	38%
4. 3 or more m; or m = or > M + FM	0	36	4	10	1.79	8%	20%
5. 1 or more C or Cn	4	19	8	19	3.70	24%	46%
6. 70%A or more	0	36	3	11	3.50	6%	22%
7. No H; or Hd:H = 2:1 or more	4	27	3	16	*7.58***	14%	40%
8. Static perseveration	0	35	0	15	*13.06**	—	30%
9. Contamination	0	33	1	16	*11.53***	2%	32%
10. Confabulation	2	28	3	17	*8.45***	10%	38%
11. Position response	2	42	0	6	4.17*[b]	4%	16%
12. Leaf or tree — 3 or 15% of content	0	43	2	5	.53	4%	10%
13. Elimination or sex	1	43	0	6	4.17*[b]	2%	14%
14. Anatomy — 3 or more or 15%; or F— anatomy	0	37	6	7	.00	12%	14%
15a. Bizarre content	3	26	1	20	*15.43***	8%	46%
b. Personal reference	0	41	1	8	*5.33**	2%	16%
c. Troubled content	4	18	4	24	*12.90***	16%	56%
16. Dysphoric shading: variations of Clob or C′	0	39	3	8	1.45	6%	16%

[a] Corrected for continuity (McNemar's Formula 86a).
[b] Less than 10 subjects.
* Significant at .05 level.
** Significant at .01 level.

As this table indicates, 7 of our proposed measures do show a significant difference between normal and disturbed groups either at the .05 or the .01 level. Two others show significance at the .05 level though there are fewer than ten subjects involved, which makes the measure of significance of questionable value. Two further signs — 1 or more C or Cn, and 70%A or more — fall just below the .05 level of significance.

This gives a total of 11 of the proposed 16 measures which either show statistically significant differences between the two groups, or differences which are close to statistical significance. This list includes the following signs which do appear to a statistically significant extent to distinguish disturbed from non-disturbed subjects:

1. Excessive manipulation.
2. F% exceeds F+%.
†3. One or more C or Cn.
†4. 70%A or more.
5. No H, or Hd: H is 2:1 or more.
6. Static perseveration.
7. Contamination.
8. Confabulation.
•9. Position response.
•10. Elimination or sex response.
11. Bizarre content; personal reference; troubled content.

† Falls just below the level of statistical significance at the .05 level.
• Fewer than 10 instances.

Of the proposed danger signs, that is of the kinds of response which we have been inclined to consider as possible trouble indicators, the following did not occur significantly more in our disturbed than in our normal population:

1. More than 25%Dd or more than one Do.

2. Three or more m.
3. Three or more leaf or tree responses.
4. Three or more anatomy responses or F– anatomy.
5. Dysphoric shading.

NUMBER OF DANGER SIGNS PRESENT

Perhaps equally important or even more important than the actual nature of disturbance signs in any given record, in detecting emotional disturbance on the part of the subject, is the *number* of such signs present in any one record.

Table III which presents the frequency of occurrence of *proposed* danger signs, and Table IV which gives the frequency of those which turned out to be *significant*, and also the cumulative percentages of distribution, show clearly the quantitative differences in number of signs present in normal and in disturbed records. Only two out of 50 normal subjects have more than three of the initially proposed danger signs, and more than half have either one or none. Conversely, only seven of the disturbed subjects have as few as three of the proposed signs. By percentages, only 4% of the normal group have four or more signs, whereas 86% of the disturbed subjects have four or more. Mean number of these signs in the normal group is 1.6; in the disturbed group, 5.0. The difference is statistically significant.

Considering merely those signs which turned out to occur with a statistically significant frequency more often in the disturbed than in the normal group, equally conspicuous differences exist. Only five normal subjects have more than one sign, whereas only two of the disturbed subjects have less than two. By percentages, only 10% of the normal group have two or more signs, whereas 90% of the disturbed have two

TABLE III—Frequency of Occurrence of Proposed Danger Signs and Cumulative Percentage Distribution

Number of subjects giving each number of signs			Cumulative percentages		
No. of Signs[a]	Normal	Disturbed	No.	Normal	Disturbed
0	6	0			
1	22	0	1 or more	88	100
2	12	0	2 or more	44	100
3	8	7	3 or more	20	100
4	1	13	4 or more	4	86
5	0	13	5 or more	2	60
6	0	7	6 or more	2	34
7	1[b]	7	7 or more	2	20
8	0	3	8 or more	0	6
Mean	1.6	5.0 (t=15.90)			

[a] This counts one score to a numbered sign even though a and b may both be present.
[b] A 7-year-old boy now seems probably disturbed, but who was included in our initial study as coming from a regular school group and judged normal by the teacher.

TABLE IV—Frequency of Occurrence of Significant Danger Signs and Cumulative Percentage Distribution

Number of subjects giving each number of signs			Cumulative percentages		
No. of Signs	Normal	Disturbed	No.	Normal	Disturbed
0	26	0			
1	19	2	1 or more	48	100
2	4	14	2 or more	10	96
3	0	19	3 or more	2	68
4	1	8	4 or more	2	30
5	0	6	5 or more	0	14
6	0	0	6 or more	0	2
7	0	1			
Mean	.6	3.1 (t = 12.88)			

or more. Mean number of these signs in the normal group is .6; in the disturbed group, 3.1. The difference is statistically significant.

Age Differences In Number of Danger Signals

Our earlier study (1), which suggests that quite different kinds of responses characterize different age levels so far as the Rorschach response is concerned, leads us to the question — do danger signals occur more, either for normal or for disturbed subjects, at some ages than at others? Tables V and VI present data on this subject. Mean numbers of danger signals at each age for both groups are given, both for all proposed danger signals, and then for only those proposed danger signals which turned out to be statistically significant.

When the entire list of proposed danger signals is considered, it appears that in our normal group, the largest number of these signs occur at 6 and at 7 years of age. For the disturbed group, high points for the occurrence of such signs are 6 years and 9 years, though at all ages except 12 years, four or more such signs occur and there is little difference in number from one age to another.

Counting only the significant signs, 7 years appears to be the high point of occurrence in our normal group, 6 years in the disturbed group, with a subsequent steady decrease in number except at 9 years when there are a few more than at 8 years.

Our earlier study suggests that of the 16 danger signs proposed here, none are found normatively except at 7 years of age, at which time C responses, confabulation, and bizarre, morbid or troubled content may occur in presumably normal subjects.

Table V—Mean Number of Proposed Danger Signs per Child — Normal and Disturbed Subjects, at Increasing Age Levels

		Number of signs		
		Normal group	Disturbed group	*t* Test
Mean, total group		1.6	5.0	15.90*
Median, total group		1.0	5.0	
Age	No. of subjects			
6 yrs.	6	2.1	6.1	
7 yrs.	7	2.8	4.8	
8 yrs.	7	1.6	4.8	
9 yrs.	7	1.5	5.3	
10 yrs.	12	1.1	5.0	
11 yrs.	8	1.1	4.4	
12 yrs.	3	1.0	4.0	

* Significant at .01 level

Table VI — Mean Number of Significant Danger Signs per Child—Normal and Disturbed Subjects, at Increasing Age Levels

		Number of signs		
		Normal group	Disturbed group	*t* Test
Mean, total group		.6	3.1	12.88*
Median, total group		.0	3.0	
Age	No. of subjects			
6 yrs.	6	.8	4.1	
7 yrs.	7	1.3	3.2	
8 yrs.	7	.7	3.0	
9 yrs.	7	.5	3.4	
10 yrs.	12	.3	2.7	
11 yrs.	8	.2	3.0	
12 yrs.	3	1.0	2.0	

* Significant at .01 level

Discussion

The "sign" approach to the diagnosis of personality difficulty as revealed by the Rorschach, though generally useful to clinicians, has not as yet been validated by experimental study. Even should research support such an approach, as the present study appears to do, most clinicians agree that it should never be used by those unskilled and inexperienced in clinical work with the Rorschach to take the place of thoughtful and painstaking evaluation of any individual protocol.

An examiner's total impression of a Rorschach response will in most cases reveal much more than any checking of a list of signs, no matter how adequate that list may be. Thus, emotional disturbance in a child subject is revealed by such things as going over and over the same blot area, by consistent dissatisfaction with responses, by extreme unevenness of responses, by presence

of various kinds of shock responses (as described by Phillips and Smith (6). Or, as Halpern (4) points out, emotionally disturbed children have innumerable ways of responding which emphasize their emotional disturbance. Some "emphasize control, conformity, and compliance, feeling a need to repress aggression and accept the dictates of authority. Thus they show marked constriction and inhibition." Others may go to the opposite extreme.

Or, with more seriously disturbed subjects, according to Halpern, "There are disturbances in identification, disorganizing anxiety, defects and distortions in reality testing beyond what could be ascribed to immaturity, and an inability to establish satisfying object relations."

Thus it should be made clear that:

1) The mere checking of a list of danger signs, no matter how adequate this list may be, should never be relied on to identify disturbance in the child subject, since it is the total impression of the record which most clinicians consider to give the most significant clues.

2) The signs suggested by us in this paper, as possible danger signs, are by no means original with us. All have been considered at one time or another or by one investigator or another to be at least suspect when found in child records.

The aim of the present paper is merely the comparison of records from a group of presumably normal subjects with records of a group of subjects comparable as to intelligence and socio-economic status, but considered to be emotionally disturbed, to determine whether or not a series of supposed danger signs occurs to a statistically significant degree more in the disturbed than in the normal group.

Such a check has heretofore been made only, to our knowledge, by Elkins, who found that such signs do not occur more in a disturbed than in a normal group. The present study indicates that a list of such signs, if carefully selected, does distinguish the two groups — occurring more in the disturbed than in the normal subjects.

SUMMARY

Rorschach records of fifty boys between the ages of 6 and 12 years, seen by the clinical service of the Gesell Institute and classified as emotionally disturbed were matched with records of fifty "normal" research cases, for age, sex, intelligence and socio-economic status.

These pairs of records were compared for presence of 16 proposed "danger signals." This list of danger signals is a slight modification of a list first proposed in an earlier study (1), and recently checked by Elkins (3) who found that her non-disturbed group actually expressed a higher number of these signs than did her disturbed group, all of whom were subjects referred to a clinic as being in need of child guidance services.

Assuming that by no means all children referred to a child guidance clinic are actually "emotionally disturbed", we selected from over 500 clinical referrals seen by us in the last five years, fifty boys whom, judged on other responses than the Rorschach, we had classified as clearly emotionally disturbed.

When their Rorschach records were checked against records of fifty matched "normal" cases, it appeared that seven of sixteen proposed measures occur conspicuously more in the disturbed cases, showing a significant difference between normal and disturbed subjects either at the .05 or the .01 levels. Two other signs show significance at the .05 level though fewer than ten subjects were involved, making the measure of significance of questionable value. Two further signs fall just below the .05 level of significance.

The eleven signs which thus appeared to a conspicuously greater extent in disturbed than in non-disturbed subjects are:

1. Excessive manipulation.
2. F% exceeds F+%.

3. One or more C or Cn.
4. 70% A or more.
5. No H, or Hd: H is 2:1 or more.
6. Static perseveration.
7. Contamination.
8. Confabulation.
9. Position response.
10. Elimination or sex response.
11. Bizarre content; personal reference; troubled content.

As to the number of danger signals present per case, if we consider all sixteen signs as proposed, only two of the 50 normal cases showed more than three of these signs, and more than half had either one or none. Conversely, only seven of the disturbed subjects have as few as three of the proposed signs. Mean number of these signs in the normal group is 1.6, in the disturbed group, 5.0. This difference is statistically significant.

Considering only those signs which occur with a statistical significance more frequently in the disturbed than in the non-disturbed group, only five normal subjects have more than one sign, while only two of the disturbed subjects have less than two. Of these signs, the mean number in the normal group is .6, in the disturbed group is 3.1. This difference is statistically significant.

These danger signs occur most in normal subjects at 7 years of age, in the disturbed group at 6 years. However, an earlier study (1) suggests that C, confabulation, and morbid bizarre or troubled content may occur at 7 years of age in presumably normal subjects.

Thus, in conclusion, it appears that seven of a list of sixteen proposed danger signs do occur to a statistically significant extent, more in a group of emotionally disturbed subjects than in a group of normal subjects matched for age, sex, I.Q. and socio-economic status. Four others occur just below a statistically significant level.

However, it should be noted that in our opinion a sign list, though useful in detecting emotional disturbance as reflected in the Rorschach response, should not be relied on as the sole measure of disturbance. The total configuration of the response must always be considered.

References

1. Ames, Louise B., Learned, Janet, Métraux, Ruth W. and Walker, R. N. *Child Rorschach responses: developmental trends from two to ten years.* New York: Hoeber-Harper, 1952.
2. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.*, 1950, *14*, 31-38.
3. Elkins, Elise. The diagnostic validity of the Ames "Danger Signals". *J. consult. Psychol.*, 1958, *22*, 281-287.
4. Halpern, Florence. *A clinical approach to children's Rorschachs.* New York: Grune & Stratton, 1953.
5. Institute of Child Welfare, University of Minnesota. *The Minnesota Scale for Paternal Occupations.* Minnesota. No date. (Pamphlet.)
6. Phillips, L. and Smith, J. G. *Rorschach interpretation: advanced technique.* New York: Grune & Stratton, 1953.
7. Piotrowski, Z. A. *Perceptanalysis: A fundamentally reworked, expanded, and systematized Rorschach method.* New York: Macmillan, 1957.

Received January 16, 1959

Sex Identification at Different Ages Using the Draw-A-Person Test

R. L. Butler and F. L. Marcuse
State University of Washington

The evidence bearing on both the incidence and significance of the sex of the first figure drawn, when Ss are told to draw a person, is confused and inconclusive. Brown and Tolor in a review of the literature bearing on this question have stated that "testing of a careful representative sample of each age group throughout early and later childhood is very much needed in order to investigate developmental changes or differences" (1, p. 207). Several studies (2, 4, 5) offer evidence to support the hypothesis that while the great majority of normal adult males draw their own sex first, it is quite common for normal adult females to draw the opposite sex first. These studies would suggest that approximately 85 per cent of males draw their own sex first while the corresponding figure for females is in the neighborhood of 60 per cent.

The assumption is made that boys and girls will also show this same sex difference in their drawings. Such an assumption tends to be *a priori*. In a study by Weider and Noller (6), it was reported that of 73 boys and 80 girls, between the ages of 8 and 10, of normal intelligence and from upper middle and lower socio-economic groups, 74 per cent of the boys and 97 per cent of the girls drew self-sex figures first. This is contrary to what occurs in adults. In another study Jolles (3) attempted to provide developmental information on children's drawings. This investigator was concerned, among other problems, with the question as to which sex children of grade school age tend to draw first. For this purpose he collected drawings, using the H-T-P test, from 8,500 children of ages 5-12, whose grade placement ranged from kindergarten to sixth grade. These children were located in five towns in Illinois. He found that of a sample of 1282 boys, of different ages, 85 per cent drew the male figure first and that of a sample of 1278 girls, of different ages, 80 per cent drew the female figure first.

Both the findings of Weider and Noller (6) and the results of Jolles (3) suggest that school age children do not order the sequence of sexes as do adult groups. Both studies however pose certain questions. Was the work of Weider and Noller representative?[1] Was their total number of Ss adequate? In Jolles study, did the fact that the classroom teacher administered the test have any effect? On what basis was the sample of approximately thirty per cent (2560) selected from the original 8500 drawings? Did his procedure of combining different ages conceal developmental changes? The problems posed by such questions were the subject matter for the present investigation.

Procedure

All students who were enrolled in the Pullman, Washington schools and who were present on the days when the drawings were obtained, served as Ss for this study. Drawings from 810 boys and 734 girls from 5 to 18 years of age and from kindergarten to twelfth grade were obtained. A total of 1444 drawings representing 83.3 per cent of all pupils attending Pullman schools were secured (a measles epidemic had lowered the number somewhat). In addition to this figure of 1444, 100 drawings, 49 girls and 51 boys, were procured from a private kindergarten.[2]

[1] This study was done in partial fulfillment of the requirements for the degree of Master of Science in Psychology at the State University of Washington.

[2] The 100 private kindergarten pupils may not be a representative sample. However a public kindergarten does not exist in Pullman.

TABLE I—Errors of Nine Judges Asked to Rate Drawings of Five to Eight Year Olds as Male, Female, or Ambiguous

Age		Judge 1	2	3	4	5	6	7	8	9
5	Male (N=25)	1	1	0	3	0	1	2	4	2
N=43	Female (N=18)	4	9	0	2	5	0	4	2	2
	Ambiguous	17	8	29	11	17	28	12	18	16
	Total Error	22	18	29	16	22	28	18	24	20
	% Error	52	42	67	37	52	65	42	56	47
6	Male (N=51)	6	7							
N=124	Female (N=73)	2	3							
	Ambiguous	13	11							
	Total Error	21	21							
	% Error	17	17							
7	Male (N=77)	2	4							
N=147	Female (N=70)	7	1							
	Ambiguous	9	7							
	Total Error	18	12							
	% Error	12	8							
8	Male (N=81)	5	0							
N=146	Female (N=65)	1	3							
	Ambiguous	4	6							
	Total Error	10	9							
	% Error	7	6							

Ss were tested in a group with the teacher out of the room. Each class, consisting of approximately 25 Ss, was given ten minutes to complete a single drawing. The following instructions,[3] the same for Ss at all ages, were read to each group by the E. "On the sheet of paper you now have, I want you to draw somebody. You may draw the person any way you want. I am not interested in how well you draw, I just want to know how you would try to draw somebody. Please do this yourself. Do not look at anybody else's paper. Go ahead and draw one person." All questions were answered with "You may do it any way you like—do whatever you think best." At the end of the ten minute period the E collected each drawing and at the kindergarten and elementary levels asked each S the sex of the figure drawn. In addition, the age (to the nearest year), grade, and sex of the S were also recorded by the E. At the junior high level and up, the above procedure was slightly changed and Ss were told, at the end of ten minutes, to record this same information on the back of their drawing.

At each age the percentage of males drawing males, males drawing females, females drawing females, and females drawing males was determined. Tables and a curve were then derived from these data in order to provide normative figures and to see whether there was an age range at which the sexes diverged in the tendency to draw self-sex figures in response to the directions, draw a person.

Jolles (3) had reported that only one per cent of the drawings of five year olds were ambiguous as to sex, but pilot data[3] had cast doubt on this statement. Consequently nine judges were given the drawings of the five year olds with the following instructions: "Please separate these drawings into a male pile and a female pile. If the sex of a figure is ambiguous or indeterminate to you, place it in a third pile." Two of the judges also did this for ages six, seven and eight. The age level involved was unknown to the

[3] A pilot study was made in a nearby Colton, Washington school to determine whether younger Ss would have any difficulty understanding the instructions and whether the time period (10 minutes) alloted was adequate. Both were found to be satisfactory. Only 1 of 1545 drawings had to be discarded.

judges and the two judges who made evaluations for ages five through eight were presented the drawings of the different age groups in random order.

RESULTS

Table I shows the errors of the nine judges who were asked to rate the drawings of the five year olds as male, female, or ambiguous (indeterminate). Since the sex of the drawn figure was known, judgments of male when the figure drawn was intended to be a female, female when the figure drawn was intended to be a male, or the judgment of ambiguous were all considered errors. For example at age five, judge four (Table I) had designated 3 of the male drawings as female, 2 of the female drawings as male, 11 of the drawings as ambiguous and thus had a total of 16 errors (37%). Judge three found 29 of the five year old drawings to be ambiguous and thus made a total of 29 errors (67%). The mean percentages of error of the nine judges in this study, for the drawings of the five year old Ss, was 51. This is not in accord with the one percent figure for this age reported by Jolles (3). The results of the two judges who also classified the drawings of the six, seven and eight year olds, showed a decrease in the number of errors as the age of the Ss increased, although even at age eight the sex of the figure drawn was not always obvious.

Tables II and III show the distribution of drawings, according to the sex of the figure drawn, for males and females of 5 to 18 years of age. These facts and the fact that males and females tend to diverge at a certain age range on the drawing of self-sex figures are shown in Figure 1. A two-tailed test of significance showed that the probability was .002 that the differences obtained in the two curves of Figure 1, would occur by chance.

DISCUSSION

The data reported show the importance of having young Ss, especially below the age of eight, designate the sex of the figure drawn. Results show that the number of incorrect judgments concerning the sex of the figure drawn decreases as the age of a S increases.

In order to reduce the number of uncontrolled factors, the E, instructions, materials, and length of testing period were kept constant for all Ss. The findings of this study, with Ss above eight years of age, would support the results obtained with adult groups, namely that while both sexes tend to draw their own sex first on the DAP, males tend to draw significantly more self-sex figures than females. Below the age of eight, the tendency to draw self-sex or opposite-sex figures does not differentiate between boys and girls. This study then may be said

TABLE II—Sex of First Figure Drawn by 810 Males 5 to 18 Years of Age

Age	Male		Female	
	No.	Percent	No.	Percent
5 (N=26)	18	69.2	8	30.8
6 (N=51)	38	74.5	13	25.5
7 (N=77)	59	76.6	18	23.4
8 (N=81)	71	87.7	10	12.3
9 (N=81)	70	86.4	11	13.6
10 (N=82)	75	91.5	7	8.5
11 (N=73)	70	95.9	3	4.1
12 (N=66)	56	84.9	10	15.1
13 (N=79)	75	94.9	4	5.1
14 (N=51)	49	96.1	2	3.9
15 (N=48)	47	97.9	1	2.1
16 (N=39)	38	97.4	1	2.6
17 (N=39)	38	97.4	1	2.6
18 (N=17)	16	94.1	1	5.9

TABLE III—Sex of First Figure Drawn by 734 Females 5 to 18 Years of Age

Age	Male		Female	
	No.	Percent	No.	Percent
5 (N=17)	7	41.2	10	58.8
6 (N=73)	9	12.3	64	87.7
7 (N=70)	18	25.7	52	74.3
8 (N=65)	15	23.1	50	76.9
9 (N=74)	21	28.4	53	71.6
10 (N=76)	20	26.3	56	73.7
11 (N=70)	16	22.9	54	77.1
12 (N=51)	14	27.5	37	72.5
13 (N=77)	23	29.6	54	70.4
14 (N=42)	11	26.2	31	73.8
15 (N=49)	12	24.5	37	75.5
16 (N=36)	12	33.3	24	66.7
17 (N=23)	13	56.5	10	43.5
18 (N=11)	7	63.6	4	36.4

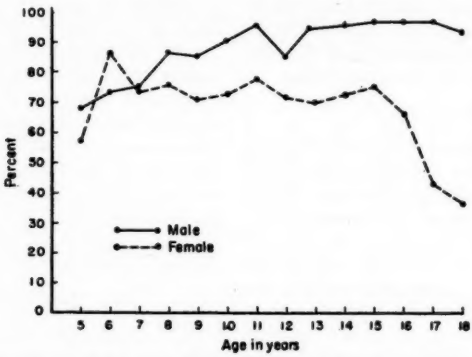


FIGURE 1. Percent of Male and Female Subjects Drawing Self-Sex Figures From Ages 5-18 Inclusive.

to yield results which both support and refute the idea that children behave like adults in their response to instructions for the DAP test. Whether the performance of children on the DAP is similar to or different from the performance of adults depends on age. The over-all results are interpreted as being related to the "dominance" (or incipient awareness of the dominance) of the male sex in the American culture as reflected in its attitudes, mores, customs, and institutions.

The present findings do not, as may be thought, have any implications for the latency period. According to Freud, the S in this period may show either reaction-formation or sublimation, in which case he might draw either the opposite or the self-sex figure. Therefore the assumption of a latency period is not open to scientific testing.

SUMMARY

The purpose of this study was to ascertain whether children, adolescents, and young adults, respond to the DAP as do adult groups. Some 810 boys and 734 girls between the ages of 5 and 18 were used as Ss. The results indicate that boys and girls above the age of eight do respond to the DAP like adult groups, but below the age of eight the tendency to draw self-sex or opposite-sex figures appears equally as strong in girls as in boys. The results are interpreted as having no bearing on the latency period but as reflecting a growing awareness among females of the dominance of the male sex in contemporary western civilization.

REFERENCES

1. Brown, D. G., and Tolor, A. Human figure drawings as indicators of sexual identification and inversion. *Percept. mot. Skills*, Monogr. Suppl. No. 3, 1957, 199-211.

2. Feather, D. B. An exploratory study in the use of figure drawings in a group situation. *J. soc. Psychol.*, 1953, *37*, 163-170.

3. Jolles, I. A study of the validity of some hypotheses for the qualitative interpretation of the H-T-P for children of elementary school age. 1. Sexual identification. *J. clin. Psychol.*, 1952, *8*, 113-118.

4. Mainord, Florence. A note on the use of figure drawings in the diagnoses of sexual inversion. *J. clin. Psychol.*, 1953, *9*, 188-189.

5. Starr, S. and Marcuse, F. L. Reliability in the 'Draw-a-person' test. *J. proj. Tech.* 1959, 23, 83-85.

6. Weider, A., and Noller, P. A. Objective studies of children's drawings of human figures. 1. Sexual awareness and socio-economic level. *J. clin. Psychol.*, 1950, *6*, 319-325.

Received October 15, 1958.

The Effect of Examiner Status Upon Rorschach Performance [1]

FRANCES A. CAMPBELL AND PAUL B. FIDDLEMAN
University of North Carolina

Several studies and theoretical analyses (2, 3, 7, 9, 10, 11, 12, 13, 14) indicate that situational variables may have an important influence upon Rorschach performance. The status of the examiner is one such variable not yet studied experimentally. The present research is directed primarily toward an experimental analysis of the influence of E's status, as perceived by S, upon S's Rorschach performance.

With the idea that more meaningful results might be obtained if the subjects were pre-selected on the basis of some psychologically relevant variable, the differential effects of need-achievement (hereafter referred to as n-ach) upon Rorschach performance under various status conditions were investigated. The n-ach scale of the Edwards Personal Preference Schedule (4) (hereafter referred to as EPPS) was used to measure this personality dimension.

An E's status might affect an S's Rorschach responses in a variety of ways. For exploratory work three measures were chosen as dependent variables most likely to be affected: the number of responses given to the cards, the number of whole responses given, and the range of content or number of content categories used (hereafter referred to as R, W, and NCC, respectively). Three predictions were made:

1. R, W, and NCC scores will be greater under high status conditions than under low status conditions.

2. R, W, and NCC scores will be greater for Ss with high scores on n-ach than for Ss with low scores on n-ach.

3. R, W, and NCC scores will be greatest under high status conditions with high n-ach Ss and smallest under low status conditions with low n-ach Ss.

PROCEDURE

Experimental Design

The experimental design is shown in Table I. Each of two Es tested equal numbers of Ss under each of the status sets, and their Ss were equally divided between high and low n-achievers.

TABLE I—Experimental Design

	High n-ach		Low n-ach	
	E1	E2	E1	E2
High status	7Ss	7Ss	7Ss	7Ss
Undefined status	7Ss	7Ss	7Ss	7Ss
Low status	7Ss	7Ss	7Ss	7Ss

Subjects

Ss were male undergraduates at the University of North Carolina, naive with respect to the Rorschach. They were selected from a pool of 430 Ss to whom the EPPS had been administered. The full range of n-ach scores (0-26) was represented in the group, with a mean of 15.1 and a S. D. of 4.0. Cutting scores (established at 10 and 20) included the upper 19% and lower 18% of the group, giving 157 persons from whom the 84 Ss were randomly drawn and assigned to the experimental groups.

Examiners

The examiners were a male and a female graduate student, the same age, with nearly identical training and ex-

[1] This work was completed as part of the requirements for the M. A. degree. Data collection was facilitated by support provided by a grant from the United States Public Health Service (Grant M-1027, E. Earl Baughman, principal investigator). The authors are indebted to Dr. Baughman, who directed the work, and to Drs. John J. Mellinger, Samuel L. Guskin, W. Grant Dahlstrom and Shephard Liverant.

perience in the use of the Rorschach.

Status Sets

The status sets were induced primarily through the use of short typewritten paragraphs describing the purpose of the study and the persons conducting it. These paragraphs were identical for the three status sets, except for the way in which E and the purpose of the study were defined. In the high status set E was defined as a research scientist, in the low status set as a student working on a term paper, and in the undefined status set Es were referred to as "individuals." The Es dressed in a manner appropriate to the status set: white lab coats for high status, casual street dress for undefined status, and, in the case of the female E, "bobby socks" for low status. Es also introduced themselves as Dr., Mr., or using their first name according to the set. Otherwise, standard procedures were maintained throughout, using bare offices.

Testing Procedure

The testing procedure included standard greeting and introduction of E, reading of the status setting paragraph by S while E left the room briefly, then the administration of the Rorschach. Beck's (1) suggested instructions were used, with S seated with his back to E. Encouragement was given if S rejected or gave only one response to Card I, never after that. Finally, Form A of the Thurstone Test of Mental Alertness (15) (hereafter referred to as TMA) was given to each S to check for possible differences in intellectual ability between the groups.

Scoring

The Es scored their protocols independently for productivity, location, and content. Scoring reliability was determined by having each E re-score twenty of the other E's records and computing reliability coefficients (Pearson product moment) for each category used. These r's ranged from +.97 to +.99.

Results

There were no significant differences between the groups with respect to intelligence. The relationships between the dependent variables (R, W, NCC) and intelligence were also analyzed, but no significant correlations were found.

The means and S. D.'s for R, W, and NCC scores obtained under each experimental condition are shown in Table II. The results of the analysis of variance for R are presented in Table III.

The analysis of variance for R shows that the number of responses given by the Ss varied as a function of the status setting used, while n-ach and the Es did not significantly affect the variance. T-tests revealed differences between the mean R in High and Undefined and between High and Low settings significant at the .01 level, but no difference between Undefined and Low settings. Ss gave significantly more responses under High status conditions than they did under the other two sets.

Correlation coefficients revealed that little of the variance in W could be predicted from R, therefore no control for R was necessitated. A logarithmic transformation of W scores was made since the variance in one cell exceeded its mean. Analysis of variance then showed that none of the independent variables significantly influenced W.

The high correlation between R and NCC (average Pearson product moment $r = +.79$) made the use of analysis of co-variance necessary in order to show how NCC varied independently of R. Two of the F ratios were significant: the interaction between n-ach and Es ($p = .05$) and the interaction between status and the Es ($p = .05$). No one of the independent variables alone, however, had a statistically significant influence upon NCC.

Discussion

The major finding of this research is that the way in which E is defined

TABLE II—Means and Standard Deviations for R, W, and NCC Obtained by Two Es from High and Low n-achievers under High, Undefined, and Low Status Settings

Rorschach Score	Status Setting		High n-ach			Low n-ach			High and Low n-ach Combined		
			E1 (N= 7)	E2 (N= 7)	Both Es (N= 14)	E1 (N= 7)	E2 (N= 7)	Both Es (N= 14)	E1 (N= 14)	E2 (N= 14)	Both Es (N= 28)
R	High	$\bar{X}$	39.9	29.0	34.4	37.3	43.1	40.2	38.6	36.1	37.3
		S.D.	18.4	14.4	16.7	13.0	21.0	17.1	15.3	18.8	16.9
	Undefined	$\bar{X}$	27.4	31.3	29.4	21.3	26.7	24.0	24.4	29.0	26.7
		S.D.	12.0	13.0	12.2	8.8	20.8	15.6	10.6	16.8	14.0
	Low	$\bar{X}$	24.9	25.0	24.9	36.7	23.1	29.9	30.8	24.1	27.4
		S.D.	6.0	17.0	12.3	20.5	11.7	17.5	15.8	14.1	15.1
	All sets	$\bar{X}$	30.7	28.4	29.6	31.8	31.0	31.4	31.2	29.7	30.5
		S.D.	14.1	14.4	14.0	16.1	19.6	17.7	14.0	17.0	15.9
W	High	$\bar{X}$	13.3	9.6	11.4	9.4	9.6	9.5	11.4	9.6	10.5
		S.D.	16.8	3.3	11.8	5.7	5.1	5.2	12.3	4.1	9.0
	Undefined	$\bar{X}$	8.7	10.6	9.7	9.0	8.1	8.6	8.9	9.4	9.1
		S.D.	1.9	7.7	5.5	4.6	4.3	4.3	3.4	6.2	4.9
	Low	$\bar{X}$	9.4	10.1	9.8	12.1	7.4	9.8	10.8	8.8	9.8
		S.D.	2.8	3.4	3.0	6.8	3.5	5.7	5.2	3.6	4.5
	All sets	$\bar{X}$	10.5	10.1	10.3	10.2	8.4	9.3	10.4	9.2	9.8
		S.D.	9.6	5.0	7.6	5.7	4.2	5.0	7.8	4.7	6.4
NCC	High	$\bar{X}$	11.3	8.7	10.0	12.7	9.1	10.9	12.0	8.9	10.5
		S.D.	3.2	4.0	3.7	4.5	4.2	4.6	3.8	3.9	4.1
	Undefined	$\bar{X}$	9.3	11.7	10.4	8.0	8.6	8.3	8.6	10.1	9.4
		S.D.	3.5	3.8	3.7	2.6	5.7	4.3	2.9	4.9	4.1
	Low	$\bar{X}$	8.1	7.9	8.0	11.9	7.6	9.7	10.0	7.7	8.9
		S.D.	1.7	3.2	2.5	3.9	2.8	3.9	3.4	2.9	3.3
	All sets	$\bar{X}$	9.5	9.4	9.5	10.9	8.4	9.6	10.2	8.9	9.6
		S.D.	3.0	3.9	5.0	4.1	4.2	4.3	3.6	4.1	3.9

TABLE III—Analysis of Variance for R

Source	df	Variance Estimate	F	p
Status	2	987.94	4.15	$.01<p<.05$
n-ach	1	68.76	<1	
Examiners	1	48.76	<1	
Interaction status x n-ach	2	541.45	2.27	$.10<p<.20$
Interaction status x Es	2	230.73	<1	
Interaction n-ach x Es	1	12.19	<1	
Interaction status x n-ach x Es	2	405.08	1.70	$.10<p<.20$
Error	72	238.14		
All sources	83	21,064.95		

to S may affect the number of responses that S gives to the Rorschach blots. Thus the significance of the interpersonal context within which the testing takes place is underscored. A possible source of variation in results from one research report to another is also suggested. Since R is significantly related to most other Rorschach variables (5, 6), the implications of this finding are even more far reaching. There is in addition the suggestion that certain interactional effects involving the examiner may influence

the range of content that S produces.

Interpersonal theory suggests that an individual's performance is not unrelated to the human situation within which it occurs. Also it seems likely that most Ss will want to do their best when working for or with a person of some significance. The nature of the Rorschach affords S no certain means of evaluating his performance, but the amount or quantity of response is oftentimes, in our culture, one means of estimating how well one has done. The Ss in this research reacted to a person with high status by increasing the number of responses emitted.

There were no differences in productivity, however, between the undefined and low status sets. Numerous remarks by Ss in the undefined context suggested that they perceived E to be another student collecting thesis data, hence the similarity of results under the two conditions may reflect only their subjective identity to most Ss. It appears to make no difference whether E is specifically defined as a student or whether the situation is handled in such a way as to make this the natural interpretation, without the words being spoken.

The finding that Ss who differ in n-ach do not differ in R was unexpected. Before this finding can be interpreted more information concerning the validity of the EPPS scale as a measure of n-ach will have to be gathered. To some extent, the results of this research question the construct validity of this scale.

The failure of W and NCC to reflect changes in E's status might best be attributed to the variables themselves. Others have found W to be a relatively stable score despite attempted experimental manipulations (8). As for NCC, a mere counting of content categories may have obscured qualitative changes in content which did perhaps occur.

Conclusion

The Rorschach was administered to 84 Ss (high and low n-achievers) by two Es under three status sets. The Ss gave significantly more responses when E was defined as having high status than when his status was undefined or when he was defined as having status equal to that of S. Status did not affect the number of content categories used, however, nor the number of W responses. No significant effects due to n-ach or Es were found.

References

1. Beck, S. J., *Rorschach's test I. Basic processes.* New York: Grune and Stratton, 1950.
2. Calden, G. and Cohen, L. B. The relationship of ego-involvement and test definition to Rorschach test performance. *J. proj. Tech.*, 1956, *17*, 300-311.
3. Cox, F. N. and Sarason, S. B. Test anxiety and Rorschach performance. *J. abnorm. soc. Psychol.*, 1954, *49*, 371-377.
4. Edwards, A. L., The Edwards Personal Preference Schedule. New York: The Psychological Corporation, 1953.
5. Eichler, R. M. Some comments on the controlling of differences in responses on the Rorschach test. *Psychol. Bull.*, 1951, *48*, 257-259.
6. Fiske, D. W. and Baughman, E. E. The relationship between Rorschach scoring categories and total number of responses. *J. abnorm. soc. Psychol.*, 1953, *48*, 25-32.
7. Henry, E. M. and Rotter, J. B. Situational influences on Rorschach responses. *J. consult. Psychol.*, 1956, *20*, 457-462.
8. Hutt, L. M. *et al.* The effect of varied experimental sets upon Rorschach test performance. *Rorschach Res. Exch.*, 1950, *14*, 181-187.
9. Leventhal, H. Effects of perceptual training on the Rorschach. *J. consult. Psychol.*, 1956, *20*, 93-98.
10. Lord, E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, 64 (Whole No. 316).
11. Phares, E. J. and Rotter, J. B. An effect of the situation on psychological testing. *J. consult. Psychol.*, 1956, *20*, 291-293.
12. Sanders, R. and Cleveland, S. The relationship between certain examiner personality variables and subject's Rorschach scores. *J. proj. Tech.*, 1953, *17*, 34-50.
13. Sarason, S. B., *The clinical interaction.* New York: Harper and Brothers, 1954.
14. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *Psychiatry*, 1945, *8*, 419-448.
15. Thurstone, L. L. and Thurstone, T. G., Thurstone Test of Mental Alertness. Chicago: Science Research Associates, Inc., 1952.

Received October 1, 1958.

The Perceptual Organization TAT Score: Number, Order, and Frequency of Components

RICHARD H. DANA
University of Nevada

Concurrent validation studies of the TAT have proposed three simple, objective scores which employ stimulus-properties of the cards (3, 5, 6, 7). These scoring categories, (a) Perceptual Organization (PO), (b) Perceptual Range (PR), (c) Perceptual Personalization (PP), were developed from a rationale for objective scoring (2) and a statement of personality theory (1). Two scoring categories, PO and PR, had approximately equivalent predictive efficiency. As a consequence further research will focus on separate categories.

PO DEVELOPMENT AND VALIDATION

PO reflects one basic aspect of test behavior, the approach to the situation. PO measures the *S*'s ability to follow the directions to "tell a story". There are seven components which epitomize the usual directions: (a) card description, (b) present behavior, (c) past events, (d) future events, (e) feeling, (f) thought, (g) outcome.

The 390 *Ss* used in the development of PO and for the present studies were equally drawn from rigorously defined normal, neurotic, and psychotic populations, 240 males and 150 females.

Two kinds of scorer reliability were determined. Scoring category reliability with randomly drawn male and female samples, from original and cross-validation groups, was 93 and 94 per cent of agreement. The reliability of scoring PO components ranged from 89 to 97 per cent. There were no differences in scoring by psychologists or clerks.

Concurrent validation studies yielded prediction scores which successfully differentiated between 68 and 83 per cent of *Ss* in each group from other groups. Sex differences are minimal. Length of TAT story and intelligence of *S* are not significant variables. PO thus has considerable value, at least for prediction of traditional clinical labels. In a sense PO scores provide an index of degrees of psychiatric illness. Details are contained in a manual summarizing development and validation research (8).

PROBLEM

Previous research has been propadeutic to validation of PO with predictions of personality characteristics as external criteria. These predictions have primary importance in the attempt to increase the probability of "correct" clinical statements from TAT data. Pilot research using the order and frequency of PO components suggested relationships to personality variables and to psychopathology. Prior to any examination of specific hypotheses, it was necessary to evaluate the stimulus-value of the TAT cards with respect to number, order, and frequency of PO components elicited. The salient question with respect to PO is: How similar are the stimulus-values of the TAT cards for number, order, and frequency of PO components for normal and clinical *Ss*?

METHOD AND RESULTS

PO stimulus-value of TAT cards: The numbers of PO components elicited by each card are roughly equivalent. Evidence for this accrues from product-moment correlations between PO scores for each card with PO scores for all cards. A mean correlation (z transformation) of .70 was obtained with a range from .60 (card 16) to .78 (card 14). Since a short-form TAT set is used throughout (4), only those pertinent

correlations will be reported: Card 2=.68; card 3GF=.76; card 4=.69; card 6GF=.64; card 7GF=.74 (N=67 normal, female college *Ss*). Table I shows the PO intercorrelations for these five TAT cards.

TABLE I—PO Product-Moment Inter-Correlations for 67 Normal Female *Ss* on Five Cards

Card	3	4	6	7
2	.62	.40	.50	.53
3		.61	.57	.58
4			.58	.57
6				.60

Note: all correlations are significant at <.001 level.

There was a .91 correlation between the mean PO score for the five cards and the mean PO score derived from all cards. Thus, each card elicits PO scores which are close approximations of (a) the mean PO score for all twenty cards, (b) scores of the other four cards in the short-form set (for females).

Order of PO components: The order of PO components does not differ either from card to card or from group to group. Table II presents intercorrelations for mean order with normal female *Ss*. Nine of ten *rho* correlations are significant at the <.01 level. Table III presents similar comparisons between mean orders of occurrence for PO components of normal and clinical groups.

There is a significant relationship between order of occurrence for all three groups. The order of PO components, irrespective of diagnostic group, proceeds from card description to either present behavior or past events, then to feeling, thought, future events and outcome.

Frequency of PO components: Table IV presents *rho* correlations between frequencies of each PO component for the three groups.

TABLE II—Rank-difference Inter-Correlations of the Mean Order of PO Components for 67 Normal Female *Ss*.

Card	3	4	6	7
2	.89	.79*	.89	.96
3		.96	1.00	.96
4			.96	.89
6				.96

Note: all correlations are significant at <.01 level except * which is significant at the <.05 level.

TABLE III—Rank-difference Correlations Between Mean Orders of Occurrence of PO Components for Normal, Neurotic, and Psychotic Groups

Group	Neurotic	Psychotic
Normal	.89	.86
Neurotic		.96

Note: all correlations are significant at <.01 level.

TABLE IV—Rank-difference Correlations Between Frequencies of Each PO Component For Three Groups

Group	Neurotic	Psychotic
Normal	.54	.50
Neurotic		.90*

* significant at <.01 level.

The frequencies of occurrence for PO components are not significantly related for normal and clinical groups. The neurotic and psychotic groups, however, have highly similar frequencies.

TABLE V—Rank-difference Inter-Correlations Between Cards for Frequency of PO Components in Each Group

	Card			
Group	3	4	6	7
Normal				
2	.00	.42	.53	.58
3		.60	.21	.58
4			.88**	.85**
6				.78*
Neurotic				
2	.29	.87**	.84*	.98**
3		.49	.60	.38
4			.96**	.81*
6				.89**
Psychotic				
2	.44	.71*	.86**	.93**
3		.76*	.53	.38
4			.85**	.76*
6				.96**

* significant at <.05 level.
** significant at <.01 level.

Rank-difference correlations were computed between cards and between groups for frequency (Tables V, VI). Inspection of Table V indicates that frequencies for some cards are significantly related; frequencies for some cards are not significantly related. With ten intercorrelations for each group, three for normals, six for neurotics, and seven for psychotics are significant. Thus, the PO stimulus-value, with respect to frequency, is much more similar for clinical groups than for a normal group. In other words, *psychopathology effectively reduces the objective differences between TAT cards.* When intercorrelations between groups for frequency on each card are inspected, the simple inference of differences between normal and clinical groups is complicated (Table VI). It appears that some cards for some groups have similar frequencies. Although differences between groups do obtain when frequencies of scores for all five cards are used, accurate diagnosis cannot be made on this basis. It is congruent with these results to assume that *personality dimensions may be distinct from diagnostic categories and that discrepancies in results may be attributable to this.* A general hypothesis accruing from this analysis is that frequency of occurrence of PO components is related to personality data.

TABLE VI—Rank-difference Inter-Correlations Between Groups for Frequency of PO Components on Each Card

	Group	
Card 2	Neurotic	Psychotic
Normal	.76*	.58
Neurotic		.81*
Card 3GF		
Normal	.16	.38
Neurotic		.87**
Card 4		
Normal	.83*	.71
Neurotic		.96**
Card 6GF		
Normal	.92**	.79*
Neurotic		.85**
Card 7GF		
Normal	.42	.35
Neurotic		.96**

* significant at $<.05$ level.
** significant at $<.01$ level.

SUMMARY AND CONCLUSIONS

1. All TAT cards are approximately equivalent stimuli. It makes no difference which cards are used.

2. The number of TAT cards used is important with five cards providing data equivalent to that from the entire set.

3. The order of PO components is not related to the specific TAT card or to diagnostic group. A standard order is preserved for all groups on all cards.

4. Frequency of occurrence of PO components is not related to card stimulus-value *per se* but to diagnosis and hypothetically to personality variables.

These results suggest the need for (a) large scale correlational attempts to isolate personality variables related to specific PO components, (b) testing of well defined groups which are high and low on previously isolated personality variables to determine the strength of relationships to PO components, (c) the development of specific hypotheses from the above results which predict the direction and magnitude of relationships between PO components and independent external criteria. This would constitute cross-validation. Research is in process on the first two aspects of this proposed framework for PO validation.

REFERENCES

1. Dana, R. H. Personality Orientation: an organizational focus for current research. *J. Psychol.*, 1954, *37*, 139-150.

2. Dana, R. H. The objectification of projective techniques. *Psychol. Rep.*, 1955, *1*, 93-102.

3. Dana, R. H. Clinical diagnosis and objective TAT scoring. *J. abnorm. soc. Psychol.*, 1955, *50*, 19-25.

4. Dana, R. H. Selection of abbreviated TAT sets. *J. clin. Psychol.*, 1956, *12*, 36-40.

5. Dana, R. H. Cross validation of objective TAT scoring. *J. consult. Psychol.*, 1956, *20*, 33-37.

6. Dana, R. H. An application of objective

TAT scoring. *J. proj. Tech.*, 1956, *20*, 159-163.
7. Dana, R. H. Norms for three aspects of TAT behavior. *J. gen. Psychol.*, 1957, *57*, 83-89.
8. Dana, R. H. Proposal for objective scoring of the T.A.T. *Percept. Mot. Skills.*, 1959, *9*, 24-34 Monogr. Suppl. 1.

Received October 21, 1958

The Relationship Between Rorschach Responses and Attitudes Toward Parents

CYNTHIA ENGEL

There has been a trend in Rorschach interpretation recently toward more emphasis on content analysis, generally based on psychoanalytic theory. One of the early references to content analysis appears in Bochner and Halpern (2). In discussing cards IV and VII these authors write: "Card IV — the heavy male figure may suggest the father or authority in general. This may be pleasant or unpleasant. It's dark quality and overwhelming characteristics are particularly disturbing to those for whom parental authority is still an unresolved problem. Card VII — the two female faces and even female figures (in reverse position dancing girls), as well as the generally soft light quality, frequently with maternal implications." Their hypotheses concerning cards four and seven are based on clinical experience and they present no further evidence in support of their hypotheses. Since the original introduction of the idea that cards IV and VII represent the father and mother figures respectively, many clinicians interpret responses to these cards as attitudes felt toward the parent figures.

The first systematic investigation to determine the extent to which the properties of cards IV and VII actually do elicit responses relative to "father" and "mother" was done by Meer and Singer (8). The hypothesis they tested was that normal adults will designate card IV as the "father card" and card VII as the "mother card" more frequently than would be attributed to chance. The subjects they used were 50 male college students. While their results were statistically significant, in actual numbers only 12 of the 50 subjects tested chose card IV as the "father card," and only ten chose card VII as the "mother card."

Rosen (11) repeated the procedures of Meer and Singer, using as subjects 180 university psychology students, and obtained comparable results.

Charen's study (3) is the most recent in this area. He administered the Rorschach, and at the conclusion asked his subjects to "pick out the card which is most like your father," and "pick out the card that is most like your mother," and they were asked the reasons for their choices. With but few exceptions the questions were interpreted by the patients as meaning "which *response* that you gave to the Rorschach reminds you of your parent." Charen tested over 50 adults in this manner. Both sexes were tested; all were white and from various socio-economic classes except relief groups. His conclusion is that there is a tendency to "use all cards in such a manner that no distinction between cards IV and VII and the other eight cards could be made."

Thus, according to Charen, if card VII is always presumed to represent the mother figure and card IV the father figure for all subjects, gross errors would result in a percentage of cases.

PROBLEM

Since the previously cited studies have been done with adults, it was felt that it would be of interest if a similar study were done with children. In addition to determining the cards chosen by children as "mother cards" and "father cards", certain other problems were also undertaken. The questions investigated in the present study were:

First, do children choose card IV as the "father card" and card VII as the "mother card" more frequently than they choose any other cards.

Second, since not all children choose cards IV and VII, is their attitude toward card VII the same as toward the card chosen as the "mother card", and

is their attitude, whether positive or negative, toward card IV the same as their attitude toward the card chosen as the "father card." If the attitudes toward the cards chosen as the "mother" and "father" cards and toward cards IV and VII are respectively the same, then if a card other than IV and VII are chosen there might still be some justification for interpreting responses to cards IV and VII as relevant to attitudes to mother and father.

Subjects

The 30 subjects used in this study were volunteers recruited from the author's classes and classes of her colleagues.[1] All subjects were boys between the ages of 11 and 12. They varied in school grade from five to seven. The children were all from middle and upper middle class homes and of the Jewish religion.

Procedure

Each subject was administered the Rorschach according to the method as outlined by Klopfer (6). At the completion of the standard Rorschach test the subject was asked to choose two cards: "One that in your imagination makes you think of your father, and one that in your imagination makes you think of your mother."[2]

After choosing the two cards the child was asked to explain why he chose them. All the subjects were able to choose two cards and explain the reasons for their choice.

All the Rorschach responses were scored as indicating positive or negative attitudes. If it was not clear during the administration whether a response was positive or negative, then during the "Testing of the limits" the subject was questioned and encouraged to talk a little more about the animal or leaf or whatever else was the subject of his response. As a result each response became scorable as either positive or negative.

Results

In Table I we have presented the frequency with which each card was chosen as a "father card" and "mother card" by the 30 children who participated. It is apparent that there is a great deal of variety in the children's choices. Although there is no strong trend toward choosing cards IV and VII respectively as the "father" and "mother" cards, card III was chosen a great deal. The two most frequent choices for a "mother card" were cards III and VII. The two most frequent choices for a "father card" were cards III and IV. Since card III is the most potent human figure in the set of cards, is there a different set of identification, diffuse and undifferentiated, in the pre-adolescent as contrasted with the adult?

A comparison was made between the attitudes expressed toward card IV and those cards chosen as the "father card" and the attitudes expressed toward card VII and those cards chosen as the "mother card." It was felt that if those attitudes are similar, then there would be justification for this type of interpretation as "mother card" and "father card." A comparison was also made between the cards chosen as mother and father cards and card X as a control card. The agreement between card VII

Table I—Choice of Mother and Father Cards

Card No.	Number of times chosen as mother card	Number of times chosen as father card
1	2	2
2	2	0
3	8	6
4	1	6
5	0	2
6	4	3
7	6	2
8	2	1
9	4	4
10	1	4

[1] The author wishes to thank the officials of the East-Rockaway Jewish Center for permission to conduct the study, and her colleagues for their help in recruiting the boys who participated.

[2] Based on a pretest which had been undertaken previously, this was found to be the most easily understood form of posing the question to the boys.

and the chosen mother card is 56.6%, and the agreement between card X and the chosen "mother card" is 40%; the difference between these two proportions is not significant at the 5% level. Agreement between card IV and the chosen "father card" is 50%, and the agreement between card X as the control and the chosen "father card" is 50%. Thus there is no greater agreement between the attitudes expressed toward card VII and those expressed toward the chosen "mother card" than those expressed toward card X and the chosen "mother card."

Neither is there any greater agreement between the attitudes expressed toward card IV and the chosen "father card" than there is between card X and the chosen "father card." Since the cards children choose as mother and father cards vary greatly, and there is no greater agreement between the attitudes expressed toward card IV and VII and the attitudes expressed toward the chosen "father" and "mother" cards respectively, and the attitudes expressed toward card X as the control card, there seems to be no justification for using cards IV and VII as the "father card" and "mother card" for purposes of interpretation.

To test whether the different Rorschach cards actually do elicit different attitudes depending on stimulus material, a correlation was computed between the percent of responses to cards 1, 3, 5, 7, and 9 which were positive and the percent of positive responses to cards 2, 4, 6, 8, and 10. The correlation is .20, not significant at the 5% level. This indicates that the Rorschach cards do elicit different attitudes, but these differences in attitudes do not in any way lend support to the theory that certain cards elicit responses that in some way are related to the mother and father figures.

Thus the two questions posed at the outset of this study have been answered. These findings fail completely to support the theory that in terms of content analysis, card IV is the "father card" and card VII is the "mother card." The results of this study together with the findings of Charen clearly indicate that to interpret blindly reactions to card IV as reactions to the father figure and reactions to card VII as reactions to the mother figure may lead to gross errors of interpretation.

However one interesting point was raised as a result of the children's choices of mother and father cards. The boys picked card III frequently for both the mother and father card. This suggests that the identification may be different for these pre-adolescents than it is for adults. To follow this through, the author is now testing girls between the ages of nine and eleven to see whether they also exhibit this diffuse undifferentiated process of identification by choosing card III frequently as the father and mother card.

SUMMARY

The study was undertaken to see whether children choose cards IV and VII as the "father" and "mother" cards, and since not all children will choose these cards, are their attitudes towards the cards chosen as "father" and "mother" cards similar to their attitudes towards cards IV and VII.

The results obtained clearly indicate that when children choose a "mother" and "father" card there is little difference between cards IV and VII and the remaining eight cards in regard to frequency of choice. This study provides no support for the theory that cards IV and VII usually represent the father and mother figures respectively. Interpretations based on this theory will be in error a sizeable proportion of the time. It was found, however, that card III was chosen frequently for both the "mother" and "father" card. The author is now doing a follow up to see whether girls between the ages of nine and eleven will exhibit this reaction too.

REFERENCES

1. Beck, S. J. *Rorschach's test,* Vol. 1 & 2. New York: Grune and Stratton, 1944-45.

2. Bochner, Ruth and Halpern, Florence. *The clinical application of the Rorschach test.*

New York: Grune and Stratton, 1942.

3. Charen, Sol. Pitfalls in interpretation of parental symbolism in Rorschach cards IV and VII. *J. consult. Psychol.*, 1957, *21*, 52-56.

4. Edwards, A. *Experimental design in psychological research.* New York: Rinehart, 1954.

5. Freud, S. *The interpretation of dreams.* New York: Modern Libr., 1950.

6. Klopfer, B., Ainsworth, Mary, Klopfer, W., and Holt, R. *Developments in the Rorschach technique,* Vol. 1. New York: World Book, 1954.

7. Lindner, R. M. Content analysis in Rorschach work, *Rorsch. Res. Exch.*, 1946, *10*, 121-129.

8. Meer, B. and Singer, J. L. A note on the "father" and "mother" cards in the Rorschach inkblots, *J. consult. Psychol.*, 1950, *14*, 482-484.

9. Peatman, J. *Descriptive and sampling statistics.* New York: Harper, 1947.

10. Rorschach, H. *Psychodiagnostics.* Berne: Huber, 1951.

11. Rosen, E. Symbolic meanings in the Rorschach cards: a statistical study. *J. clin. Psychol.*, 1951, *7*, 239-244.

12. Schafer, R. *Psychoanalytic interpretations in Rorschach testing.* New York: Grune & Stratton, 1954.

Received December 29, 1958

Types of Pre-Cognitive Defensive Organization in a Tachistoscopic Experiment

ULF KRAGH
Stockholm, Sweden

INTRODUCTION

The starting-point for the experiments reported here has been the theory of defense mechanisms formulated by Sigmund and Anna Freud and others, and the actual-genetic model of perception-personality (6, 7, 8, 13, 15). S. Freud (4) has given a scheme for the dynamics underlying defense mechanisms. Very roughly this scheme comprises three links: threat/danger, anxiety, and defense against anxiety. We have transformed this scheme into experimental procedure, equally encompassing three links: *subliminal perceptual threat, (subliminal) anxiety, and defensive response.*

The actual-genetic model of perception-personality has been conceived in close interaction with an experimental technique for the study of pre-cognitive organization and of the development of percepts (6, 7, 8 and 12, 13, 14, 15; cf. also 2 and 16). We are not entering in detail upon the theoretical foundations or the various technical aspects; they have been laid down elsewhere. Though the experiments on "defensive pre-cognitive organization" presented here have been worked out within the frame of the actual-genetic *(act)* theory of perception and personality, interpretations of the findings do not necessarily presuppose the acceptance of the model as a whole.

During the last decade the concept of "perceptual defense" has been the object of continuous discussion. Even if there is affinity of approach with respect to techniques there are, however, great differences between the analysis of pre-cognitive *organization* given here, and the analysis of "perceptual defense" mainly aiming at establishing the frequency of raised *thresholds* for correct recognition.

METHOD AND DEFINITIONS

The technique consists in the successive exposure of a TAT-like picture by the tachistoscope, exposure time being progressively increased.

The tachistoscope is made of a projector with a 100 Watt bulb and a camera shutter. A piece of paper is placed around the projector tube so as to screen off the experimenter's manipulations from the Ss. The pictures are projected onto a transparent screen 25"x14". The screen is placed between the projector and the Ss at the distance where the projected (and focused) picture just covers the screen. No fixation point is utilized. The illumination of the room is regulated so that the picture is barely recognized clearly and distinctly at unlimited exposure time. Under the standard conditions and with the standard diapositives employed this equals scale values .4-2.0 on a Weston Master Universal Exposure Meter placed at the distance of 5 m. in front of the screen, and directed towards the projector.

There are 12 exposures of each picture, in msecs: 20, 40, 40, 100, 100, 200, 200, 200, 500, 500, 500, 500.

The stimulus employed here is the TAT picture number 1 (boy with a violin), the head and shoulders of a threatening and ugly male person having been inserted at the right above the boy. The experiments are conducted with groups of not more than about 40 Ss. The first row of Ss is placed at the distance of about 6 m. from the screen. (Mentally abnormal Ss are tested individually.)

The Ss are instructed to make a drawing of what they have seen on the screen without paying attention to whether their impression is correct or not. For each exposure they have to make a drawing in a square. If they feel unable to make any kind of drawing, they are also allowed to make markings instead of a drawing. In addition to the drawings (markings) a written comment is also required so as to make quite clear what each S has seen, and at what place on the screen. The Ss are told to work as fast as they can without neglecting the thoroughness of description. The time interval between the exposures is determined by the

time required by the slowest S for reporting. It has rarely been more than about 3 minutes.

In all 260 aviation cadets took part in the experiments; their age varies between 18 and 21 years. They were all subjected to a preliminary screening procedure, medical and psychological. None has an I.Q. much below the average, and visual acuity is normal. Their educational level varies between elementary school and college (Swedish Gymnasium).

Definitions. *Actualgenesis* denotes method as well as ("pre-cognitive") data in their serial arrangement. The concept *phase* is defined as the sum total of data registered in connection with one exposure. In the experiments referred to here, the first phase corresponds to the exposure at which S has first seen something to which he attributes a meaning, and the last phase that exposure at which he has seen the picture "correctly". P-phases we term all those phases which precede the last phase.

Defensive Organization and Defense Mechanisms

The analysis of certain well definable types of pre-cognitive organization will be related to the frame of psychoanalytic defense mechanisms for the following reasons:

1. Specific types of pre-cognitive organization have been found to characterize compulsion neurotic patients and have been associated with the defense mechanism of isolation (6 pp. 47, 191-204 and 288; 14).

2. In a small group of patients suffering from anxiety hysteria, the following phenomena were observed. Exposing the TAT picture nr 6GF by the tachistoscope in the manner described above, an anxiety-laden black spot gradually emerged at the place of the *man* in the picture. But at the very exposure where the man's head was on the verge of becoming recognized by the (female) patients, they perceived instead an ugly, threatening *devil's mask*. For three reasons this particular pre-cognitive structure was tentatively identified by referring to the defense mechanism of *repression*. Since S. Freud (4, pp. 63-65 and 155-158), repression is considered to be the main mechanism operating in hysteria. The threat presented by the picture was a "sexual" one, and the transformation of the visual field occurred at that perceptual level where the heterosexual relation was barely becoming recognized by the Ss (cf. 1, pp. 148-150). Finally the structural characteristics of "rigidity" (possibly connected with hysterical paralysis) also made it plausible to refer to the mechanism of repression (9).

3. Granted that no sharp borderlines separate the "normal" personality from the "abnormal" one as regards the working of defense mechanism (the occurrence of "defensive" pre-cognitive organization), analogous mechanisms (types of pre-cognitive organization) should be found in normal Ss with "abnormally strong" stimuli as in abnormal Ss with "weak" or "moderately strong" stimuli. If a threat is very strong, it seems reasonable to assume that even ordinarily rather well-adapted Ss are likely to take refuge in defensive measures (cf. 10 p. 582). Starting with the experiences with compulsion neurotic and hysteric patients mentioned, the normal secondary person in the (TAT) picture was thus replaced by an abnormally threatening one in order to evoke analogous abnormal (defensive) responses in normal Ss.

Personality descriptions and observations of behavior of aviation cadets who failed in primary training were made by officers of the training staff. These descriptions, too, lend support to an identification of defensive organization with defense mechanisms. They have not, however, been formalized for a rating procedure.

4. We have already pointed at the close correspondence between Freud's scheme and the scheme of our experiments. In many cases we find manifest anxiety reported by the S at one or more exposures, the anxiety referring to a black or dark part of the visual field; at the next exposure this organization is replaced by another less stimulus-conformable one, the anxiety (dis-

pleasure) disappearing more or less completely.

Utilizing Freud's terminology we say that in our experiments, secondary anxiety is induced, "in which something occurs in the id which activates a danger-situation for the ego and induces the latter to give the anxiety signal for inhibitions to take place" (4 p. 113). Following Freud we assume that the reaction to an internal danger-situation (which is created in our experiments by the short exposures of threatening pictures) is analogous to the reaction to an external danger-situation.

In the experiments neither the threat nor the anxiety must necessarily become conscious in order to provoke a defensive reaction. This is effected by using the projection of the ego instead of inflicting pain or displeasure directly on the Ss' ego (as in many experiments on perceptual defense). On the other hand there are Ss in whom the anxiety is manifest but defensive reactions are minimal, resulting in reality-testing.

5. There are obvious structural correspondences between types of pre-cognitive organization in our experiments on one hand and types of pre-conscious organization, i.e. defense mechanisms, on the other (1, 3, 4). When, for example, the existence of a threat is explicitly denied by the S at one exposure, we may refer to the defense mechanism of denial. If the threat is seen to be directed from the hero toward the secondary person, we may refer to identification with the aggressor.

6. The technique of free association, which has been the paramount tool of psychoanalysis, involves the reconstruction of pre-conscious material, neatly corresponding with the reconstruction of pre-cognitive levels (or of P-phases) in the act experiment. The concept of defense mechanism was mainly derived from data obtained by means of the method of free association.

We cannot, therefore, agree with G. Klein (5, pp. 349-350) who maintains that it is "pointless to expect direct analogues of 'repression' in percep-

NOTE. Markings made by the Ss: pe=person, m=man, k=woman, g=boy, f=girl, +=sympathetic, −=unsympathetic. Numbers represent age.

FIG. 1 — Repression (4) and isolation (2)

"A sleeping boy with a violin on his knees. A flower pot on the table."
(S nr 1, exposure 9)

FIG. 2—Repression (2) and reaction-formation

"A boy playing with a puppy."
(S nr 2, exposure 3)

FIG. 3—Isolation (1 and 4)
"An odd old fellow to the right."
(S nr 5, exposure 8)

FIG. 4—Isolation (double 1)
"To the right the boy was seen in a mirror."
(S nr 5, exposure 9)

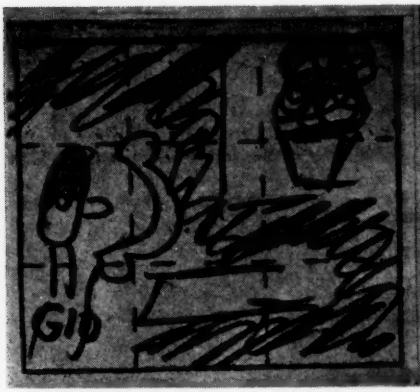


FIG. 5—Isolation (double 1) and repression (4)
"A flower in a window."
(S nr 5, exposure 10)

FIG. 6—Reaction-formation
"A girl is sitting in the other bank, writing."
(S nr 6, exposure 12)

tion" due to "the differences of method which both concepts reflect: perceptual methods are radically unlike those which gave rise to the classical defense mechanisms".

7. In neuroses one of the main characteristics of the defense mechanisms is considered to be that of distorting and counteracting reality-testing. Correspondingly we assume that in the development of percepts in the direction of reality (or stimulus), it is a sign of the activity of defense mechanisms when the threat (on the TAT-like picture) is not adequately dealt with in terms of reality, in particular at long exposure times (in late phases, on late perceptual levels), and when a preceding relatively correct organization is replaced by a less correct one.

CLASSIFICATION OF PRE-COGNITIVE DEFENSIVE ORGANIZATION

The following classification of pre-

FIG. 7—Identification with the aggressor
"A soldier during fight. Critical situation," (S nr 7, exposure 4)

FIG. 8—Repression (5)
"A woman sitting at table, being served." (S nr 8, exposure 8)

FIG. 9—No defensive type of organization
"An unsympathetic man appeared. Threatens the woman with a pistol." (S nr 8, exposure 12)

FIG. 10—Turning against the self (1)
"Maybe patient on the operating-table during stomach operation. A male assistant, and a female doctor?" (S nr 8, exposure 13)

cognitive defensive organization has been undertaken on the basis of drawings and reports of 260 aviation cadets (mainly those who failed in primary training).

Hero denotes the person who is seen by the S at the place of the main person in the picture. *Secondary* denotes the person who is seen at the place of the secondary person in the picture. At the beginning of many series only one person is seen; in such cases we also term this person hero on the condition that he is not situated at the place of the secondary person in the picture.

The figures illustrate some types of

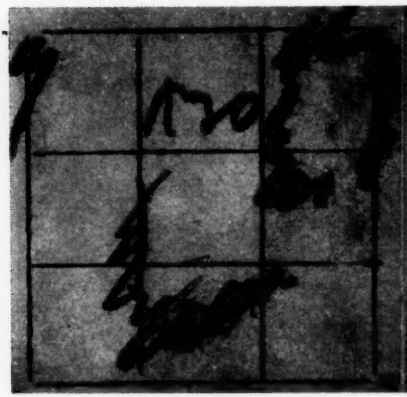


FIG. 11—Isolation (5)
"A woman at her lace-making table." (S nr 8, exposure 15)

FIG. 12—Isolation (5) and Turning against the self (1)
"Somebody lying on the floor in the foreground." (S nr 9, exposure 13)

defensive organization. They are drawings made by seven of the Ss. It should be underlined that all drawings refer to one single picture motif.

Repression

The hero or the secondary are living but not human beings, or lifeless objects.

(1) The hero, or the secondary, or both are rigid, inanimate beings. Examples: sculpture, bust, statue.
(2) The hero or the secondary is an animal. (Fig. 2)
(3) The secondary is a "mask", or "disguised".
Mixed type (repression and isolation, 1): "a man disguised as a ghost".
(4) The hero or the secondary is an object. (Fig. 5)
Mixed type (repression and isolation, 2): lamp, window.
(5) The hero or the secondary is only a part of the body (except for the face). (Fig. 8)

Isolation

The hero and the secondary are separated or isolated from each other. The main person, or the secondary person, or the threat are not seen. The pre-cognitive levels are excluded (isolated) and there is only one single quasi-objective level of cognition.

(1) The hero and the secondary are isolated from each other in the visual field. (Figs. 3, 4 and 5) Examples: (a) The hero is separated from the secondary by a line, a dark surface, or an object. (b) The hero turns away from the secondary or *vice versa*. The hero puts his fingers into his ears or holds his hands before his eyes. This type of organization may be looked upon as a mixture of isolation and denial. (c) The persons are referred to different frames; they belong to different levels of reality or the secondary is a reflected image of the hero or they "do not have any connection with each other" or one of them is in perspective in space, the other not.
Mixed type (isolation and repression): the secondary, but not the hero, is a picture; (isolation and denial): "they are not conscious of each other's presence".
(2) The secondary is a white spot. (Fig. 11)
Mixed type (isolation and repression): lighted lamp, bright window.

(3) There is no secondary at all through the whole series, not even in the form of a threatening dark part of the field.
The secondary person is not seen, though the main person has appeared more or less in detail during many previous exposures.

(4) There is a secondary who is, however, neither threatening nor even unsympathetic. (Fig. 3) These structures represent mixed types of isolation on one hand and of denial or reaction formation on the other hand.

(5) After one or many persons have been seen at one or many previous exposures, the persons disappear at one or many subsequent exposures: phase discontinuity. (Fig. 11)
Phase discontinuity at one exposure and in one series only may be due to blinking.
Discontinuity may affect (a) the whole visual field (zero phase), (b) hero only (hero discontinuity), or (c) only the secondary (secondary discontinuity), or it may affect a dark, threatening spot at the secondary person's place (threat discontinuity).
Mixed type (isolation, 1 and 5): the whole picture is seen as "unreal", "like a dream or a phantasy".

(6) There are no changes of contents all through the series.
The contents may be essentially correct from the first exposure and afterward with respect to schematic characteristics (the hero's age and sex, etc.).

Denial

The threat is explicity denied in the S's report.

Mixed type (denial and isolation): The secondary is "a threat which has not been drawn correctly," or the whole picture is said to represent "a puppet show." In these cases the threat is partly denied, and partly isolated as incorrect.

Reaction formation (Fig. 6)

The threat is turned into its opposite.

Examples: The secondary is "nice", "protective" or "friendly" to the hero.

In some cases it would be difficult to make any sharp distinction between denial and reaction formation, e.g., when the hero smiles or when the emotional atmosphere is experienced as "pleasant".

Identification with the aggressor (Fig. 7)

The role of the main person is reversed, the hero becoming the aggressor. Mixed type (identification with the aggressor and reaction formation): the hero protects the secondary.

Turning against the self (Figs. 10 and 12)

The hero or the hero's instrument is damaged or worthless, or the hero's instrument is a threat to the hero.

(1) The hero or his instrument (corresponding with, e.g., the violin in the picture TAT nr 1) is a sick, bleeding, crying, or dead person, or a person committing suicide.

(2) The hero's instrument is something damaged or broken: a ragged toy, a broken tool. In some cases the whole picture is seen as "an explosion," "the Guernica."

(3) The hero's instrument is something dark *and* threatening.

Defensive organization which is not directly identifiable as a classical defense mechanism: Identification with a female role.

(1) The hero is a female person and the hero's instrument or tool a rounded or/and hollow object (ball, marble, etc.).

(2) The hero is a woman or a girl, and the hero's instrument a baby-doll in the woman's (girl's) arms.

DISCUSSION

It has proved possible to identify, in a group of aviation cadets, types of precognitive defensive organization previously found in compulsion neurotic

and (anxiety) hysteric patients, and also to associate preliminarily other types of pre-cognitive organization with the defense mechanisms of denial, reaction formation, identification with the aggressor, and turning against the self. Results obtained from predictions of failure and success in primary training of aviation cadets on the basis of ratings of defensive pre-cognitive organization, seem to substantiate the assumption that such organization may be referred to the classical defense mechanisms.

It should be stressed that this analysis of pre-cognitive defensive organization is a preliminary one. Much research work with groups of abnormal Ss, and with single cases diagnosed by psychoanalysts, will be required in order to state more exactly which type of pre-cognitive organization to refer to what specific defense mechanism.

The operational definitions of defensive organization in this article refer to one single exposure (phase) only, or to the sequence of two exposures (phases). We have refrained from discussing the questions concerning the strength of the (single) defensive organization, the significance of the pre-cognitive level, the various combinations of pre-cognitive organization, and the types of sequence. We are working through these areas and at the same time trying to establish rating procedures for personnel selection and for clinical diagnosis (the PDO-test, 8). It is supposed that diagnosis of neuroses *via* functions (in our experiments = pre-cognitive defensive organization) should be preferred to that *via* symptoms (1 p. 194).

Summary

In experiments with tachistoscopic exposure of TAT-like pictures, exposure time being repeatedly increased until nearly correct recognition, a threat to the hero was introduced. Types of defensive (visual) organization below the threshold of correct recognition are preliminarily classified and referred to the defense mechanisms of repression, isolation, denial, reaction formation, identification with the aggressor, and turning against the self. Other types of responses, too, are marked off as supposedly defensive. A number of facts and arguments for associating pre-cognitive defensive structures with the classical defense mechanisms are discussed.

References

1. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

2. Flavell, J. H. and Draguns, J. A microgenetic approach to perception and thought. *Psychol. Bull.* 1957, *54*, 197-217.

3. Freud, Anna. *The ego and the mechanisms of defence.* New York: Internat. Univ. Press, 1957.

4. Freud, S. *Inhibitions, symptoms and anxiety.* London: Hogarth, 1949.

5. Klein, G. S. The personal world through perception. In R. R. Blake and G. W. Ramsey (eds.), *Perception: an approach to personality.* New York, 1951.

6. Kragh, U. *The actual-genetic model of perception-personality.* Lund: C. W. K. Gleerup, 1955.

7. Kragh, U. Pathogenesis in dipsomania. An illustration of the actual-genetic model of perception-personality. (To be published)

8. Kragh, U. Analysis of pre-cognitive defensive organization: a new method for diagnosis and personnel selection (the "PDO-test"). (To be published)

9. Kuhn, R. Maskendeutungen im Rorschachschen Versuch. *Monatsschr. f. Psychiatr. u. Neurol.*, 1943, *107*, 1-60 and 1944, *109*, 169-256.

10. Murphy, G. *Personality: a biosocial approach to origins and structure.* New York: Harper, 1947.

11. Sells, S. B. and Holtzman, W. H. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, *49*, 485-490.

12. Smith, G. The serial mirror-drawing test. *Acta Psychol.*, 1957, *13*, 288-298.

13. Smith, G. Visual perception: an event over time. *Psychol. Rev.*, 1957, *64*, 306-313.

14. Smith, G. and Henriksson, M. Studies in the development of a percept within various contexts of perceived reality. *Acta Psychol.*, 1956, *12*, 263-281.

15. Smith, G. and Kragh, U. Do microgenetic sequences reflect life history? An illustration. *Acta Psychol.*, 1955, *4*, 504-512.

16. Werner, H. The concept of development from a comparative and organismic point of view. In *The concept of development: an issue in the study of human behavior.* Minnesota, 1957.

Received November 12, 1958
Revision received April 30, 1959

Projective Test Assessment of the Effects of Sleep Deprivation[1]

NATHENE TURK LOVELAND
and
MARGARET THALER SINGER
Walter Reed Army Institute of Research

Although many investigators (3, 5, 16) report "alterations in personality" among the most important manifest effects of prolonged sleep loss, findings from specific personality test measurements are lacking.[2] The so-called "personality changes" attributed to acute sleep loss usually refer to three features which are described as nearly universally experienced: hallucinatory experiences, irritability, and subjectively felt changes.

Pilot studies of prolonged sleep loss by the present investigators and their colleagues revealed the need for more precise operational definitions of such changes and helped us reach the following conclusions:

1. Definition of the stress stimulus (e.g., a certain number of hours of sleep deprivation) does not constitute an adequate operational definition of the stress, since it is variably experienced, reacted to, and communicated by different subjects.

2. Reactions attributed to sleep loss have not been adequately screened from reactions to the experimental setting which also is variably experienced. For example, to some subjects, the Walter Reed experimental setting was a bonanza of free snacks and attention for which sleep loss was a small price to pay; for some subjects it was a frustratingly restrictive interlude, where deprivation of privacy, freedom of movement, decent clothing and women, was a heavy addition to the load of sleep deprivation; for others it was primarily a challenging, competitively stimulating experience, and for still others the setting was relatively unimportant.

Various methods of handling the subjects is another aspect of the experimental setting which may affect the kind of "personality changes" found. For example, in the Walter Reed project where the subjects were handled gently and permissively, increased irritability was not clinically manifest. Similarly, the incidence and kind of hallucinatory experiences may be affected. Myths about what happens to subjects during prolonged loss are transmitted in the literature and may affect the results through both staff and subjects. Further, the interview methods used may channel what is communicated so that stereotyped interactions are set up which elicit biased answers. For example, what one examiner asks may influence a sleep-deprived subject to reply in ways that will class his experience as a hallucination whereas another questioner in another setting may elicit mere sensory disturbance reports from him. Similarly, the questioner may inadvertantly affect the incidence of hallucinations reported in the different sensoria.

3. Precise definitions of the non-

[1] We are indebted to Major Harold Williams, Chief of Clinical and Social Psychology, and Dr. Ardie Lubin, Research Psychologist, both of WRAIR, Forest Glen, Maryland, for their advice on the statistical as well as many other aspects of this study; to Dr. David McK. Rioch, Director of the Division of Neuropsychiatry, WRAIR; to Miss Elizabeth Broomhead, Chief Clinical Psychologist, National Naval Medical Center, Bethesda, Maryland, and to Dr. Marion Handlon, Staff Psychologist, Chestnut Lodge Sanatarium, Rockville, Maryland, for their blind analyses of our Rorschach records; to Mr. Jim Kincannon for his analysis of our Iflund results, and Miss Betty Shanks of WRAIR for her Devos scoring of our Rorschach records.

[2] Tyler (15) reported that group Rorschach examinations given before and after 100 hours of wakefulness showed no significant changes.

verbal, behaviorally non-overt, subjectively felt changes are especially hampered by linguistic limitations and here, just the process of describing to another person may alter the felt state.

4. Testing, as with describing subjectively felt changes, may alter the state being tested. Those who have worked with people subjected to prolonged sleep loss have been impressed, as we were, with the subject's capacity to pull himself together, alerted by the stimuli of test and tester, and to respond more or less as he did before sleep deprivation, no matter how he appeared or behaved between tests.

5. Finally, we are not sure precisely what personality tests measure, nor is it known whether or how they would be sensitive to short-term alterations of personality functioning.

This exploratory paper deals with two major types of questions:

1. What, if any, changes do projective tests show which are attributable to sleep loss?

2. From the projective test performance before sleep deprivation, can sleep deprived behavior be predicted? Of the many psychological and physiological aspects of sleep deprived behavior which might profitably be investigated in this way, for only the following two were criteria available:

a. How well will the sleep deprived subject maintain his mental and psychomotor efficiency?

b. Which sleep deprived subjects will hallucinate?

The Research Setting

In the 1957 WRAIR sleep deprivation project, a research group composed of neurophysiologists, psychiatrists and psychologists studied the effects of 98 hours of sleep deprivation on 25 enlisted men from an Army field hospital training unit who volunteered for the experiment.

The subjects reported in groups of ten for a twelve-day period. Each of the five groups was divided into five experimental and five control subjects roughly matched for intelligence. After an orientation day, there were three control days. From the beginning of the fifth day, the experimental group was kept awake for 98 hours. Following a one-day break during which the experimental group were sleeping, there were three more control or recovery days.

The 25 experimental and 25 control subjects were confined to a suite of rooms from the first orientation day through the last recovery day. Both groups were in hospital garb. The suite of rooms was equipped with materials for handicrafts and games, with television and radio sets, books and magazines and a refrigerator for snacks.

A nurse, several medical corpsmen, and members of the staff kept the men under close observation, especially during the sleep-deprivation period. Those persons and the subjects themselves helped those of the men who needed it to stay awake during their sleepiest phases by walking, playing games or talking with them. The attitudes of the subjects toward the experiment and the staff were excellent for the most part, with no one dropping out of the experiment before its completion.

Experiment Design

Four projective tests were administered to the fifty subjects on three occasions spaced as follows, so that each subject's sleep deprived performance could be evaluated in the light of (a) his own performance before sleep loss and after his "recovery" sleep, and (b) the performance of roughly matched control subjects given the same testing schedule but allowed to sleep eight hours each day:

1. Baseline test (on the third control day).

2. Sleep deprivation test (on the last experimental day after some 80 hours of wakefulness on the part of the experimental subjects).

3. Recovery test (on the third recovery day).

Exactly the same form of the four tests were readministered in the same order in the same testing room, and for the most part, by the same author. The three performances of a given subject in each of the four tests were scored at the same time to promote consistency in scoring.

The four projective tests used[3] were:

1. The Rosenzweig test (14), scored according to the directions in the manual.

2. The Iflund test (6), analyzed for shifts in order of recall.

3. The Glad test (4), scored for "interpersonal" qualities in verbalization according to the directions in Leary's manual (8). The scoring was done blind, and the order in scoring the three performances of each man was randomized as was the order in scoring the experimental and control Ss.

4. The Rorschach test (13), analyzed on the basis of

a. Scores:

(1) Thirty-eight different standard scores,[4] ratios, and percentages were calculated. Having first agreed on a common scoring system and spot-checked for reliability of scoring, each author scored the records of the subjects she had tested.

(2) The affective tone of the content was quantified by the Devos method (2).

b. Clinical interpretations:

(1) By the authors:

(a) A-posteriori interpretive analysis of differences between sleep deprived and non-sleep deprived performance.

(b) Prediction of which sleep deprived subjects would hallucinate.

(c) Subsequent a posteriori qualitative description[11] of differences between the baseline Rorschach performance of the subjects independently rated as "hallucinators" and "non-hallucinators."

(2) By two consultant Rorschach experts:[1] (whose interpretations were "blind" in that they had no contact with the experiment other than the Rorschach protocols, but were given the description of the research setting, so to that extent were working in context).

(a) Selection of the sleep deprived protocol from pairs of records of the experimental subjects. One consultant was given 12, the other 13, randomly selected protocols of the 25 experimental subjects in pairs and told that one of each pair was given before, and the other after 80 hours of sleep deprivation. Obvious retest clues such as "I remember this from last time," and reaction time were deleted.

(b) Selection of the experimental pairs from the control pairs. Each consultant was given the experimental pairs she had not seen before plus an equal number of control pairs, was told these would be presented to her in random order. Nothing but obvious sleep deprived remarks were deleted

[3] The Rorschach test, in which the authors were most interested and experienced, was given more exhaustive treatment than the other three projective tests. Even the experimental design was more appropriate for the Rorschach than for the other tests. For example, the Iflund test as used here could not be submitted to an expert for blind differentiation of experimental test from experimental retest, or experimental test and retest from control test and retest, on the basis of sleep deprivation effects, because one could tell at a glance which was which from the number of cards recalled at first showing.

[4] The 38 separate standard Rorschach scores analyzed were t_c, $t_{non\text{-}c}$, av.t, #R (number of responses), number of rejections, M. FM, m, FC, CF + C, c (all shading), #P, P%, W%, Dd%, S% (all space responses), F%, F + %, extended F + %, H%, A%, Anat.%, % other stereotype, EB% (ratio of M + Sum C to R), eb% (ratio of FM + m + c to R), M% (ratio of M to M + Sum C), M'% (ratio of M to M + FM), M''% (ratio of M to M + FM + m), FM'% (FM + m to FM + m + c), C% (Sum C to Sum C + Sum c), %R to last 3 cards, W'% (ratio of A + H to A + H + Ad + Hd), A + H — Ad — Hd, M + Sum C, M — Sum C, FM + m + c, FM + m, —c.

since the order of administration within each pair was given.

(c) Prediction from the subject's baseline record of how well he would stand up under the stress of prolonged sleep deprivation in terms of a rating scale[5] hereinafter referred to as the *h-rating*.

(d) Prediction of which subjects would hallucinate when sleep deprived. Here the consultants performed two operations:

1. First they were asked to predict whether or not the subject would hallucinate under the stress of sleep deprivation and to state the reasons for their predictions. No other directions were given.

2. Two months later each was again given the baseline Rorschach protocols of the 16 men for whom independent hallucination ratings were available (see Hallucination Criterion below) along with both the a-posteriori descriptive summary[6] of the differences between the baseline Rorschach records of the subjects classified as hallucinators and those not so classified, and a checklist[7] derived therefrom, then asked to:

(a) Rate these 16 subjects on each of the seven items on the checklist.

(b) Label each subject diagnostically. In addition. they were asked to state whether "hysteric traits" or "obsessive traits" were more prominent, regardless of the diagnosis.

(c) Predict which subjects would hallucinate after prolonged sleep loss, basing their judgment on the authors' descriptive summary. They were given no information as to the success of their prior predictions.

[5] Place each subject on a six point scale as to how well he will probably stand up under the stress of prolonged sleep deprivation, rating as "6" the subjects you think most likely to hold their own or improve on their pre-sleep deprived efficiency level in mental and psychomotor tasks, and as "1" the subjects you think most likely to fold under stress.

[6] Cited in footnote 14.

[7] Cited in footnote 15.

Criteria

1. *Hold (H) Criterion:* The following six tests were used to assess how well the subjects maintained their own mental and psychomotor efficiency level on retest because they showed significant mean decrement with sleep loss, and were administered at about the same times as the projective tests to a large enough number of the men so that each of the 25 experimental and 25 control subjects had been given at least three of the six tests.

a. An *auditory vigilance* task requiring the subject to press a lever when and only when a critical signal occurred (17).

b. A *reaction time* task requiring the subject to respond as rapidly as possible to a visual signal (17).

c. A *concept attainment* task requiring the subject to determine the properties that defined class membership (17).

d. An *adding* task requiring the subject to add pairs of single digit random numbers as fast and as accurately as possible (9).

e. An *immediate recall* task requiring the person to recall all the Iflund cards after as few re-showings of the cards as possible (9).

f. A *rhythm* task requiring the subject to press a button at various rates ranging from once per second to once every eight seconds (10).

Sleep deprivation did not have a general effect on any given subject in all these tests; it was not uncommon to find a sleep-deprived subject who improved as well as the best control subject on one of these tests, yet deteriorated drastically on another. Nevertheless, it was possible to rate the men on a six-point scale, hereinafter referred to as the *H-criterion*, ranging from those who held their own or improved in most of these tests (rated as "6"), to those who fell down on most of them (rated as "1"). Independent rating of the men in this respect by one of our statistical consultants compared reliably with the authors' ratings.

2. *Hallucination Criterion:* From tape recording of sleep-deprived psychiatric interviews plus the clinical observations of staff members, 16 of our 25 experimental subjects were rated by a staff psychiatrist and psychologist (11) on a six-point hallucination rating scale. This ranged from "no reported changes" through "illusory changes" to reported experiences where the subject labelled or interpreted the illusory change and was enough confused about, or, for a time at least, convinced of the reality of the experience to act upon it (e.g., not just that spots on the linoleum floor seemed to the subject to move, but that he thought they might be bugs and tried to stamp on and squash them). Only the latter subjects were classed as "hallucinators."

RESULTS

I. *Sleep loss effects on projective test performance.*

Our over-all conclusion from the results of repeated projective tests[8] is that prolonged sleep loss leaves personality structure essentially unchanged, but does modify how, and how much of himself, the subject communicates. The sleep deprived men changed significantly more than did the controls on retest, but like the controls, *how* they changed was a complex function of the interaction of their long term personality adjustment with the whole experimental setup, and not of the sleep loss alone.

Findings which led to these conclusions:

a. From statistical analysis of scores:[9]

(1) Only three of the more than a hundred different test scores showed changes in central tendency significant at the .05 level of confidence or better:[10]

(a) The Rorschach M% (ratio of M to M + Sum C) increased with sleep deprivation (.01 level of confidence).

(b) The sado-masochistic Iflund card was recalled earlier after prolonged sleep loss. (.01 level of confidence.)

(c) The Devos Rorschach N% (% of responses neutral in affective tone) increased on the second, sleep deprived period testing to about the same extent for both the experimental and control subjects but the experimental subjects increased their N% even more on the third, recovery period performance whereas the control subjects returned to their baseline levels.[11] (.02 level of confidence.)

Since three out of 100 changes might easily appear significant by chance, any or all three of these may be spurious results. Further, the clinicians' analyses did not support the hypothesis that these findings were attributable to sleep loss.

(2) The Rorschach test showed no cumulative over-all effect of sleep deprivation on scores.[12]

(3) Only two of the more than a

[8] It may be that the first experience with a projective test sets up a pattern of response so that subsequent use of that test material is colored by this initial experience and remembrance. An alternative explanation of the paucity of change found here, and the one we favor, is that these tests measure aspects of personality *structure*, which is fairly stable.

[9] Since we had no evidence on which to base a priori hypotheses about score changes to be expected with sleep deprivation, only two-tailed tests of significance were used.

[10] Kendall's W was used to test for trends from baseline through sleep deprivation to the recovery period.

[11] Tyler's finding (16) that "tests on the morning after recovery sleep showed no changes from pre-experimental control values other than those due to learning" was not the rule in our data. Further, clinical evidence of delayed reactions to sleep loss and of considerable individual variation in rate of recovery, reported by our staff observers, supports our impression that the recovery period performance cannot be assumed to be a "recovered" performance, thus cannot be used as a second sample of baseline performance.

[12] The nature of the tests, and our lack of a priori hypotheses as to the direction, if any, toward which sleep deprivation would move the scores, preclude an adequate assessment of the possibility that prolonged sleep loss

hundred scores showed changes from baseline to sleep deprived performance that were differentially related to their respective baseline scores:

(a) Subjects who gave less than 20 baseline responses (#R) tended to give even fewer responses after prolonged sleep loss (chi-square test, .02 level of confidence).

(b) Subjects who gave less than six popular responses (Beck P) (1) tended to give a decreased #P after prolonged sleep loss (chi-square test, .01 level of confidence).

Although no such test-retest relationships proved significant for the control subjects, again two out of 100 scores might easily show a significantly curvilinear test-retest relationship by chance. Furthermore, in most cases the change was very small, often amounting to a difference of only one R or P, so that, for a sample the size of ours, replication would be especially necessary. In the case of the #R, replication with the same results would still not constitute evidence that the change is a function of sleep deprivation, since it has been our experience that the retest #R tends normally to be curvilinearly related to the original #R, i.e., that subjects with low R tend to decrease their number of responses on retest whereas those with relatively high number of responses are more apt to hold their own or increase, contrary to the more usual homeostatic tendencies. The fact that this was found to a significant degree for the sleep deprived subjects and not for the control subjects suggests that the sleep deprivation effect, if any, was an accentuation of an already existing relationship.

(4) The experimental subjects showed a greater absolute amount of test-retest change than did the control subjects, not only, and not even especially, on test scores whose central tendencies changed most.

(a) For every one of the 38 Rorschach standard scores, the sum of the absolute amount of change was greater for the experimental than for the control subjects.

(b) For only four scores, however, was the absolute change significant (Mann-Whitney U test):

1. H% (.001 level of confidence)
2. Average reaction time (.01 level of confidence)
3. C% (Sum C to Sum C+Sum c) (.01 level of confidence)
4. Dd% (.01 level of confidence)

(c) A crude discrimination score, computed by summing these four scores was 70 per cent successful in discriminating the experimental from the control subjects when the median was used as a cutting point, and the optimal cutting point was 82 per cent correct.

b. Findings from the consultants' analyses:

(1) Neither consultant had better than chance success in discriminating between the baseline and sleep deprived performance of the experimental subjects. Both reported that the decisions were difficult, often having to be forced, because (a) in a third of the cases they found essentially no differences between the two records of the same man, and (b) when they did consider the records enough changed to suggest that something other than a brief lapse of time occurred in the intertest interval, the changes were still relatively slight, never basic, and seldom simply or consistently in a single direction.

Each consultant reported independently that she usually selected the less adequate as the sleep deprived per-

might have an over-all cumulative effect on central tendency even though so few of the separate scores changed significantly. In the case of the Rorschach test, however, we hazarded a guess that sleep loss would, if anything, increase the mean of the reaction time, the A%, the S%, other stereotypy, and the number of rejections, and lower the mean #R, the F + %, the P%, the C%, the M′%, and the W′%. Accordingly all these variables were added to form a single sum score, the Kendall W of which was not significant.

formance. We know of no established criteria for adequacy of adjustment but, to the extent that their value judgments were valid, their negative results suggest that sleep deprivation does not make for less adequate Rorschach performance.

However, the value of these negative findings as evidence that sleep deprivation does not significantly change Rorschach performance is weakened by the following methodological artifact: the order of administration was not given because the consultants knew that the retest was always the sleep deprived performance. In essence their goal was discrimination of test from retest on the basis of sleep deprivation effects; in practice they sometimes found themselves deciding which of the two records was sleep deprived on the basis of retest effects, in spite of conscientious efforts to the contrary.

To the extent that this occurred, consciously or unconsciously, their analysis in this operation became an unmeasurable mixture of a priori and a posteriori judgments about sleep loss effects. Thus even had the discriminations been highly successful, they might have had little to do with sleep deprivation. The varying accuracy of their judgments about retest effects further complicates interpretation of their negative results on this operation. The unreliability of retest clues is seldom recognized, so deceptively logical do they appear on hindsight. For example, in a few of these cases, the authors had feared that even after deletion of reaction time and obvious retest comments, the practice effect was so obvious it would give the show away, whereas the consultants reported retest bias in the opposite direction.

(2) The consultants were 68 per cent[13] correct (significant at the .05 level of confidence) in differentiating the experimental from the control records, given the baseline and retest record of each subject in pairs, with the order of administration correctly labelled. Each reported independently that she had leaned more heavily on the amount than on the kind of change, and in fact, most of their errors in judgment consisted in labelling as "experimental," control cases where they found appreciable change, and labelling as "control," experimental cases in which they found relatively little change. They reported appreciable test-retest differences significantly more often in the experimental pairs than in the control pairs, (17 experimental vs. 10 control). No significant generalizations emerged from their descriptions of what to them constituted "appreciable change" but partial clarification comes from the fact that in 8 of the 25 experimental pairs and only 3 of the control pairs they described the "appreciable change" largely in terms of what they called signs of fatigue (e.g., compared with his original performance the man became slaphappy, more careless, showed depletion or conservation of energy).

c. Authors' a posteriori analyses:

(1) More of the experimental than control records appeared appreciably changed. In one third of the Rorschach records of the experimental subjects and two thirds of the control subjects, essentially no test-retest differences appeared.

(2) For the rest, where changes did appear,

(a) In no case was there basic change in personality structure. Some subjects showed more of themselves and some less, some loosened, others tightened their controls, but it was the self-expression and not the self expressed which changed.

[13] While on the surface it appears as if the clinicians were less successful than the sum of the scores H%, t, C% and Dd% in their discrimination of experimental from control subjects, the reverse is probably the case since their judgments are a priori whereas the scores were the a posteriori discriminations of four scores selected out of 38, and it is very likely that the score-based per cent of correct clasification would be decreased on a cross validation sample.

(b) The experimental subjects changed more than did the control subjects but no kind of sleep deprived change occurred which was not also found among the retest changes of control subjects.

(c) No kind of change occurred often enough to be considered a characteristic sleep deprivation affect. Each subject seemed to be changing in his own way, which appeared to be a complex function of the interaction of his baseline personality adjustment with the whole experimental set up and not a function of sleep loss alone.

(d) No kind of change occurred significantly more often among the experimental than among the control group with the following possible exception, which at least partially described the change in self-expression or communication: while both the experimental and control subjects often showed the retest practice effect of verbalizing their perceptions in less detail but more glibly, in 7 of the experimental cases and only 3 of the control cases, retest communication seemed either more drastically curtailed and condensed than normal, or, if less curtailed, the added verbiage was more redundant than enlightening, revealing an "anything goes" attitude. The fact that, with one exception, these were the same men reported by the consultants as "fatigued" indicates reliable spotting of essentially the same quality.

(e) Only one generalization predicted the test-retest changes of all the subjects, experimental and control, better than the hypothesis that there would be essentially no change: the task-involved subjects who (either because of inner standards or "other" directed drives to please) marshalled their resources at the first challenge of the relatively unfamiliar, unstructured Rorschach test, either held their own or increased in this respect on retest. Those who were little interested, challenged, or involved originally, tended to fall even farther short of their productive capacity on retest.

Reliability was high between the way the authors rated the subjects in this respect and the consultants' h-ratings, which in turn correlated significantly with how well the man stood up on retest in a variety of nonprojective tasks (H-criterion). Further, it will be recalled that the #R, which was found to be the most important single indicator of the h-rating by all four clinicians, showed a small but significant curvilinear test-retest relationship.

(f) Those who changed most in the group as a whole, experimental as well as control, appeared in general to be the more conflict ridden, loosely defended neurotics (and not, as we had anticipated, those with the most schizophrenic-like records, who were among those showing least change). There was insufficient agreement among the authors and the two consultants to substantiate this general impression adequately. However, the psychiatrist (12) who conducted and analyzed the psychiatric interviews of these subjects reported that it was the loosely defended subjects who showed the greatest sleep deprivation effects. Further, he too concluded that personality structure was essentially unchanged by sleep deprivation when he found that the factors which differentiated the sleep-deprived from the control interviews were clearly present in the baseline Rorschach personality assessments.

(g) Reliability among the clinicians as to which subjects did and did not show essentially any change was high:

82% agreement among authors and consultants
84% agreement between consultants
68% agreement between clinicians and optimal cutting point of 4-score sum above described.

II. *Prediction Results.*

Using the Rorschach alone, it was possible to predict (a) to what extent a man would maintain performance on a wide variety of tasks during the sleep deprivation period, and (b) which men would report hallucinations and which would not during the sleep de-

privation period. However, the fact that the retest efficiency of the control subjects was predicted equally successfully by the clinicians raises the question of how much of the predicted effects are attributable to sleep loss.

Findings which led to these conclusions:

1. Prediction from the baseline Rorschach performance of how well the man would hold up (H-Criterion):

a. Only one of the projective test scores, the number of responses (#R), differentiated the subjects who stood up well under sleep deprivation from those who showed decrement in most of the six tasks of the H-Criterion significantly better than chance ($r = .42$, .05 level of confidence). The #R it will be recalled also predicted its own direction of change in the testing instrument itself.

b. The consultants' "blind" a priori predictions of how well the subjects would stand up under stress (the h-ratings) correlated with the H-Criterion significantly better than chance ($r = .43$, .05 level of confidence).

Neither of these results can be assumed to be predictions of sleep loss effects since:

(1) The consultants' h-ratings of the *control* subjects predicted their retest efficiency (H-Criterion) at the same level of confidence as for the sleep deprived subjects.

(2) One out of 100 score changes might well appear significant by chance. While the h-rating finding appears to support the #R finding, especially since the clinicians found the #R their most important single indicator of h-rating, it is supported not as a sleep deprivation effect but as a normal predictor of test-retest change. Further, though the #R did not predict the H-Criterion for the control subjects significantly better than chance, there was a tendency in that direction ($r = .23$). The fact that #R predicts the H-Criterion significantly better than chance for the experimental and not for the control subjects appears largely a function of the significantly greater absolute amount of change in the experimental subjects on retest in both the projective and nonprojective tests.

2. Hallucination predictions:

a. Only one of the projective test scores, the number of Beck popular responses (#P), successfully differeniated the subjects rated as hallucinators by other staff members (Hallucination Criterion) from those not so rated ($r = .62$, .02 level of confidence). The #P it will be recalled also successfully predicted its own direction of retest change for the sleep deprived subjects in the testing instrument itself.

Although again one out of a hundred relationships might easily appear significant by chance, it is not surprising that a relatively small #P (below average perception of things as others see them), which is one of the classic Rorschach "signs" of schizophrenia, should be a correlate of our "hallucinators." It cannot be assumed, however, that the hallucinatory experiences of sleep deprived subjects are the same phenomenon as those of schizophrenics nor that it was primarily those with schizophrenic personality traits to start with who hallucinated with sleep loss. Eleven of the 16 men for whom hallucination ratings were available were classed as hallucinators according to the criterion used here. While most (4 of the 5) subjects with schizophrenic features in their Rorschach records were classed as hallucinators, the converse did not hold; most of the "hallucinators" did not give Rorschach records, baseline or sleep deprived, with prominent schizophrenic signs. Nor were intelligence, task-involvement, or maintenance of efficiency levels, correlates of the hallucination classification; neither the ACB scores nor the h-ratings or H-rating correlated significantly with the hallucination classification.

b. The consultants classified only 56% of the subjects correctly according

to the Hallucination Criterion used here, but they were referring to the hallucinations of schizophrenics. Each reported independently that she was predicting that a subject would hallucinate if and only if schizophrenic Rorschach signs were prominent in his record. Four of the five subjects thus selected by them as probable hallucinators *were* so classed by the Hallucination Criterion and the fifth was considered a chronic denier and a probable hallucinator by our colleagues.

c. The authors, who agreed in their predictions as to whether each of the 16 subjects would or would not be classed as a hallucinator, were 90% correct in their predictions, but they were predicting not just which subjects would hallucinate under stress, which was what the consultants were asked to do, but which subjects would be likely to report or reveal hallucinations in this setting. It seemed very unlikely that such a list would include all those and only those who would have hallucinatory experiences during prolonged wakefulness in any setting. Some of the latter might defensively deny them. On the other hand, among those who do report hallucinations during prolonged wakefulness, some of the so-called "hallucinatory experiences" might easily be artifacts of interpersonal interaction; the subject has to cope not only with his disturbingly unformed, unusual sensations, associations, affect, etc., but also with the difficulty of describing them to the interviewer. Thus even without the biases and anxieties of the interviewer, some subjects might be expected to resort to the anxiety-reducing maneuver of concretizing into "hallucinations" what were merely vague sensory disturbances. The fact that past and current reports of hallucinations during sleep deprivation was a matter of common knowledge to our subjects facilitated such concretizing tendencies.

d. When they were later asked to base their hallucination predictions on the authors' a posteriori descriptive summary[14] of the difference between those classed here as hallucinators and those not so classed, one consultant was 69% correct, the other 81% correct. The two consultants agreed in 12 of the 16 cases (75%) and in 11 of the 16 cases (69%) the authors and both consultants labelled the men the same in this respect. Thus the descriptive summary, in spite of its generalities and wide use of non-specific terms, was useful, indicating that it is possible to predict the subject's hallucination classification *reliably* as well as significantly better than chance.

However, the attempt to evaluate the relative importance of the descriptive summary components as predictors and to arrive at a more precise operational definition by making a specific

[14] "Non-hallucinators" appeared to know what they were doing and gave "reality-bound" responses, the kind of banal, ordinary responses usually given for the areas the subject used. They usually gave healthy records from the point of view of controls and defenses. When they gave minus responses, they seemed less sure of these or showed other evidence of realistic critical discrimination.

"Hallucinators" showed one or more of the following:

a. Personal, idiosyncratic material, odd or unusual content, or weird percepts emerged; the basic percepts may have been banal or plus but were personally elaborated, or the elaborations were clearer than the basic percept or the person elaborated on *his* associations more than on the basic percept.

b. The subject's style of handling the test situation and his experiencing of his own ideas was an unclear event. Some sounded as if they were never quite sure what was expected, not sure but what they were just guessing all the time, unsure of their percepts or their terms; they showed imprecise use of language; their ideas trailed off.

c. The hysteric type personality was more apt to be classed as a "hallucinator" than the well-defended obsessive; i.e., not a clearly differentiated world to which he responds; a propensity for intrusive thoughts and affect to emerge into his thinking and be reacted to as if they were in the outer world. (Here not necessarily CF but openly noticeable in content, where a banal or pleasant content may end up dangerous or vice versa; incompatible affect breaks through, not an integrated part of the associative process.

checklist,[15] securing diagnostic classifications, and the rating of hysteric vs. obsessive defenses, fell far short of success. Not one of the consultant's ratings or nosological classifications showed better than chance agreement either between consultants or with the ratings of the authors on this checklist. Nor did the consultants appear to be using the checklist, since in a few cases, their predictions as to whether or not the man would hallucinate ran counter to most of their ratings on the checklist and in even more cases, their ratings on the same man were about equally divided among scores which would have put the man in the hallucinator, the non-hallucinator, and undecisive class, according to our descriptive summary. Further, their cumulative ratings for each man on these various checklist items proved neither a valid nor a reliable predictor.

CONCLUSIONS AND SUMMARY

1. Personality *structure,* as reflected in the projective tests used here (Rorschach, Rosenzweig, Iflund and Glad tests), is unaltered by acute sleep deprivation, and personality *functioning* is changed relatively little.

a. There was little test-retest change for the group as a whole, experimental as well as control, when consideration is given to the large number of scores and ratings statistically tested in this study.

b. More of the sleep deprived than control subjects changed, and changed more, on retest, but for both groups it was the style and content of communication and not the personality structure which changed.

c. Where changes were found, they appeared to be a complex function of the interaction of the subject's long term personality adjustment with the whole experimental set up and not a function of sleep loss alone.

d. The increased irritability alluded to in other sleep deprivation clinical reports was not found in this study. Our experiences suggest that some reactions atttributed to sleep loss are largely a function of the experimental setting.

2. Using the Rorschach test alone, it was possible to predict (a) how well the subject would maintain his efficiency in a variety of mental and psychomotor tasks during the sleep deprivation period, and (b) whether he would hallucinate. Although these prediction results were shown to be more convincing evidence of the predictability of behavior from personality structure than of sleep deprivation effects, the following information about the latter emerged:

a. The subjects whose baseline Rorschach performance showed high motivation, either because of inner standards or strong drives for approval, tended to maintain or raise their mental and psychomotor efficiency levels on retest; those not so motivated tended to show retest decrement in a variety of non-projective psychological tests. While this relationships prevailed equally among both the experimental and control subjects, the decrement was accentuated by sleep deprivation.

b. It was the loosely defended neurotics, whether experimental or control, who showed most test-retest change in personality functioning, but again, among the loosely defended neurotics, it was the sleep deprived subjects who showed greater incidence and degree of change.

c. The men who reported hallucinations during prolonged wakefulness were not typically those who showed mental and psychomotor decrement nor those whose personality functioning changed most, but usually fell into

[15] Rate each subject on a three-point scale as to:

1. Sureness of percepts.
2. Sureness of performance.
3. Relative sureness about plus and minus responses.
4. Per cent of unusual content.
5. Per cent of personally elaborated content.
6. Presence of intrusive thoughts and affect.
7. Strength of defenses and controls.

one of the following two groups: (1) those characterized by a loosely defended hysteric-like style of adapting and (2) those with prominent schizophrenic traits.

References

1. Beck, S. J. *Rorschach's test*, Vol. I, II, and III. New York: Grune and Stratton, 1952.

2. Devos, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, *16*, 133.

3. Edwards, A. S. Effects of loss of 100 hours of sleep. *Am. J. Physiol.*, 1941, *54*, 80-91.

4. Glad, D. D. and Shearn, C. R. An emotional projection test. *Percept. mot. Skills* (Sup. 1), 1956, *6*, 1.

5. Handbook Staff, Tufts College, Part 7 of *Handbook of human engineering data*. Medford, Mass.: Tufts College, 1949.

6. Iflund, B. Selective recall of meaningful materials as related to psychoanalytic formulations in certain psychiatric syndromes. Unpublished doctor's dissertation, Univ. of California, Berkeley, 1953.

7. Kleitman, N. *Sleep and wakefulness*. Chicago: Univ. Chicago Press, 1939.

8. Leary, T. *Interpersonal diagnosis of personality*. New York: The Ronald Press, 1957.

9. Loveland, Nathene and Thaler, Margaret. Unpublished sleep deprivation study.

10. Lubin, A. Unpublished sleep deprivation study.

11. Morris, G. and Williams, H. Unpublished sleep deprivation study.

12. Morris, G. Personal communication.

13. Rorschach, H. *Psychodiagnostics, a psychodiagnostic test based on perception*. New York: Grune and Stratton, 1942.

14. Rosenzweig, S. Rosenzweig P-F Study. Author, Western State Psychiatric Hospital, 1944.

15. Tyler, D. B. Motion sickness and physiological problems peculiar to parachutists and other airborne troops. Final report, OEMcmr-282 Part 2: Fatigue and prolonged wakefulness.

16. Tyler, D. B. Effect of amphetamine sulphate and some barbiturates on fatigue produced by prolonged wakefulness. *Am. J. Physiol.*, 1947, *150*, 253-262.

17. Williams, H. L., Lubin, A. and Goodnow, Jacqueline. Impaired performance with acute sleep loss. *Psychol. Monogr.* (in press).

Received November 9, 1958

The Structured Doll Play Test as a Projective Technique for Use With Children[1]

DAVID B. LYNN AND ROSALIE LYNN
University of Colorado School of Medicine

Despite much doll play research literature and evidence for the diagnostic usefulness of doll play procedures, no doll play technique has evolved employing simple, standardized, easily quantifiable characteristics of projective tests.

This paper describes the Structured Doll Play test (SDP)[2] which is designed specifically as a clinically useful doll play projective technique (8). The paper also describes research results suggesting the concurrent and construct validity of this instrument.

The SDP was developed because clinical experience with children suggested the need for a projective test for use with children from three through 11 more suitable than those available. The first approach in developing the SDP was to use dolls in the presentation of the Duess incomplete fables (3, 9) rather than simply administering the fables orally. Duess orally presented incomplete fables as a method of studying resistance in the psychoanalysis of children between three and 15. The Duess fables suggested some of the features that are still retained in the test, although it has been modified so often that the similarity between the doll play situations and the fables is now remote.

The SDP consists of presenting the *S* with doll-house dolls in a series of typical family and age-mate situations. Each situation is structured by telling the *S* what is going on, and he is allowed to resolve the situation through doll play. Thirty to forty-five minutes are usually required in administering the test. Children characteristically enter wholeheartedly into the play and give clinically rich responses.

Following is a list of the situations included in the SDP with brief statements of their content:

1A Crib or Bed Choice. *S* is asked to decide whether the ego-doll wishes to sleep in the bed or crib, both of which are presented in the arrangement.

1B Bottle or Cup Choice. *S* decides whether the ego-doll will drink milk from a cup or bottle.

2A Peer Choice. *S* has the ego-doll choose whether he (or she) wishes to be with a boy doll or girl doll.

2B Visitor. The unchosen doll in Situation 2A is presented as visiting the ego-doll, and *S* elaborates what happens.

3A Fighting. Two same-sex dolls are said to be fighting. *S* resolves the situation and is asked to tell what they are fighting about.

3B Misbehavior. Opposite-sex children are said to have done something wrong. *S* is asked to explain.

4A Feeding. *S* chooses and elaborates on which parent feeds the ego-doll.

4B Injury. The ego-doll is hurt. The kind of injury and its seriousness are elicited.

5A Pursuit. *S* chooses which parent is running after the ego-doll and explains the circumstances.

[1] This article is based on a paper of the same title read at the May, 1958 Rocky Mountain Psychological Association meeting in Santa Fe, New Mexico. The research findings were from Rosalie Lynn's master's thesis in the Department of Psychology, Purdue University. Acknowledgment with gratitude is made to Dr. Janet Lyon for direction of the investigation, and to Dr. William L. Sawrey, University of Colorado School of Medicine, and Mr. James A. Norton, Purdue University, for statistical consultation.

[2] Published by Test Developments, Box 8306, Denver, Colorado.

5B Reprimand. *S* chooses the parent who is to be the one scolding the ego-doll.

6. Parental Relationship. *S* tells what happens in a situation involving the parent dolls alone together.

7A Separation. The parent dolls leave the ego-doll alone for a while. *S* tells what happens.

7B Hospital. The ego-doll is said to be hospitalized. Reasons for this and consequent events are elicited.

8A Toilet. *S* chooses which parent will take the ego-doll to the toilet. The ego-doll may use either toilet or potty chair.

8B Bathing. Opposite-sex child dolls are bathing together, and *S* relates what happens.

9. Sleeping Arrangement. A double and a single bed are presented, and *S* puts the parent dolls and the ego-doll to bed.

10A Nightmare. The ego-doll is said to have a bad dream. *S* elaborates on this and is asked for content and resolution.

10B Night Fears. The ego-doll is said to be afraid of something he sees or hears during the night. *S* elaborates on this.

Most of the SDP situations also involve one or both parents, often calling for a choice on the part of *S*.

Situations 1A, Crib or Bed Choice, and 8A, Toilet, are presented below to illustrate more fully the nature of the SDP situations. In Situation 1A, *S* is told, "The little boy (or girl, the ego-doll is determined by the sex of *S*) has two places he (she) can sleep. He (she) can sleep in either the crib or the bed. You put him (her) to sleep." *S* is then asked, "Who comes in, Mommy or Daddy?" and, "Then what happens?"

In 8A, *S* is asked, "Which one takes him (her) to the toilet, Mommy or Daddy?" *S* is then asked, "Which one does Mommy (Daddy, depending on the choice) take him (her) to, the toilet or potty chair?"

A study, conducted by Lillie and Lillie (5), suggests that a structured doll play approach may have the advantage of eliciting more clinically useful material than free doll play. Using eight- through 12-year-old male subjects, they compared the effects of a structured doll play approach (though not the SDP) with free doll play using comparable materials. They found that the structured approach elicited significantly more *incident agreements* than free play. By *incident agreement* is meant any obvious parallel between responses in the doll play and a parent interview. Structured doll play also elicited significantly more identification of the *Ss* with the dolls, more hostility, and more parental discipline responses. In addition, the *Ss* tended to accept the structured play better than the free play. The material, which was better elicited by the structured procedure, would seem to be clinically useful.

The SDP uses the standard doll family (mother, father, child) rather than a duplicate of the child's own family constellation. This is done since Robinson (11) found that variation in the doll family constellation had little effect on the scored variables with the exception that identification with the dolls is greater when the child's own family is duplicated by the dolls. The kind and amount of aggression did not differ significantly. In a clinical (rather than research) setting, it is not always possible to obtain valid information concerning the family constellation necessary to duplicate the child's family with dolls. It is felt that the gain in standardized procedure and in simplification by using the standard family outweighs the loss in the *S's* identification with the dolls.

The 10 SDP situations consist of eight which are divided into two comparable sub-situations, thus making 18 situations in all. In each of these eight situations, so divided, the *S* is presented with a choice of dolls, seven

of mother or father, one of the boy or girl. Ammons and Ammons (1) conducted a study which suggests that structuring parent-choice situations might be a useful projective device. They found adequate reliability in consistency of choice, and they discovered significant age and sex differences in parent preference. The nature of the choices conformed so well to cultural patterns that some evidence for the validity of the doll play parent-choice technique can be inferred.

In the SDP, after the *S* has made a choice in a given situation, he is administered a comparable situation using the parent doll which was not chosen in the original one. In this way the *S* has equal opportunity to fantasy about both parents. For example, if the mother doll is chosen to take the ego-doll to the toilet in Situation 8A, Toilet, the father doll is introduced in Situation 8B, Bathing. There is an advantage over free doll play in this regard since, in free play, a child sometimes concentrates all his fantasy on one parent doll, ignoring the other completely.

The SDP has eliminated the elaborate setting, the doll-house rooms, the suite of furniture, etc., typical of traditional doll play procedures. All of the SDP materials can be contained in a small carrying case. The test can be administered on a small space such as the typewriter leaf of a desk. It is felt that having so few materials reduces the exploratory and tangential behavior which plagues traditional doll play procedures using elaborate materials. To reduce distraction, only the materials that are used in a given situation are placed before the *S*, the others being kept in the closed carrying case.

The use of doll furniture that is crudely representative rather than realistically detailed, is based on findings in a study by Phillips (10). Phillips found that the use of ambiguous doll figures and block-like toy furniture, constructed crudely to represent items rather than in realistic detail, reduces exploratory behavior. However, extremely ambiguous materials are not used in the SDP because the nature of the test requires the *S* to recognize the furniture (e.g. toilet or potty) and to recognize the doll figures (e.g. mother or father doll).

It is felt that a structured doll play procedure is less conducive to tangential and exploratory behavior than free doll play in that it focuses the child more closely on the clinically rewarding fantasy.

The SDP differs in many ways from the Driscoll (2) and Whitmore (4) doll play kits. The Driscoll kit claims only to furnish standard materials, a 23 x 20 inch, five-room apartment, 27 pieces of furniture, etc., but it does not furnish structured situations. The Whitmore kit offers standard materials and also suggests a number of incomplete stories which might be used with the kit. It does not include choice situations, nor is it designed as a standardized projective test as is the SDP.

SDP Research

Fortunately, the period in Clinical Psychology is virtually ended when a psychologist can originate a test and place it on the market without supporting research. Some research has been completed with the SDP and more is currently under way. The following report of two studies offers results that suggest areas of validity.

In the first study (12), there were 80 mother-child pairs from a typical sailor district in Norway.[3] There were 40 mother-child pairs in a father-absent group (21 boys and their mothers, 19 girls and their mothers). In this group the fathers were sailors and were away from home at least nine months out of the year and sometimes as long as two years or more, with only infrequent and irregular pe-

[3] The senior author collaborated with Mr. Tiller in conducting this research at the Institute for Social Research in Oslo, Norway.

riods at home. There were also 40 mother-child pairs in the control group (20 boys and their mothers, 20 girls and their mothers). The children's ages ranged from eight through nine years six months. The two groups were comparable in the teacher's rating of the child's intelligence and in the father's occupational status. Both groups were essentially middle-class, the sailors all being officers. The children were administered the SDP and their mothers a semi-structured interview.

In this study of father-absence in Norwegian sailor families, construct validity in the SDP is suggested by the fact that the findings support a theoretical formulation of the effects of father-absence. A number of predictions were explicitly stated prior to the data collection, most of which were supported by the SDP results. For example, it was predicted, and supported by SDP findings, that mothers of father-absent children would be more over-protective than mothers of father-present children, that father-absent children would respond with dependence, and that boys would be more strongly affected by father-absence than girls.

Concurrent validity is suggested by the fact that the SDP discriminated between groups, i.e. the results differentiated between the sexes, and between father-absent and father-present groups. For example, a significantly higher proportion of girls than boys chose the mother doll more often than the father doll. Interestingly, a significantly higher proportion of father-absent than father-present boys chose the father doll more often than the mother doll, and the crib rather than the bed. A significantly higher proportion of father-present boys gave aggressive doll play responses between age-mate dolls than did father-present girls. A significantly larger proportion of father-absent than father-present girls gave high dependency responses in the SDP.

The following is a report of findings concerning the SDP responses of four- and six-year-old children.

Subjects

The subjects were 49 middle-class children from a private school in Indianapolis. There were 26 four-year-olds (12 boys, 14 girls) and 23 six-year-olds (11 boys, 12 girls).

Scoring Procedures

The SDP was scored for dependency, mature-immature choices, and parent doll choice. The child's dependency score was determined by the proportion of dependent to independent responses in the SDP. A response was scored dependent if the child doll was shown to seek help, sympathy, consolation, support, attention, affection, protection, or physical closeness. It was also scored dependent if the child doll was shown as immature, e.g. by crying. Independence, on the other hand, was scored if the child doll showed self-sufficiency or self-assertion, e.g. by carrying out the major part of an action by himself. The agreement between independent ratings by two judges on this variable was 92%.

In the mature-immature coding, a mature choice was scored for bed rather than crib and also for cup rather than bottle. Conversely, crib or bottle were the two immature choices.

In the early modification of the SDP used in the present study, there were only two possible parent choices rather than the eight parent choices in the current revision.[4] In the Play Situation the *S* chose the parent doll with which the ego-doll wished to play a game. In the Construction Situation the *S* chose the parent doll from which the ego-doll wished help in building something with blocks.

Results

In the statistical treatment, analysis

[4] Dr. Bertram D. Cohen, Lafayette Clinic, Detroit, Michigan, collaborated in the development of this early modification of the SDP.

of variance tested results for dependency. Chi-square was used in all other tests of significance with the exception that Fisher's exact test was used in one analysis of the data concerning bottle-cup choice where the frequencies were especially low.

Concurrent validity in the SDP is suggested in this study by the differentiation between groups (age and sex comparisons). Construct validity is suggested by the logical relationship between these findings and developmental patterns.

Table I indicates a significant age difference for dependency scores. Table II indicates that, as expected, four-year-olds received higher dependency scores than six-year-olds. Table I also indicates a significant age-sex interaction. Table II shows no appreciable difference for dependency scores between four-year-old boys and girls, but six-year-old boys received lower dependency scores than the girls and than the four-year-old boys.

In Table III, there was a trend, although not statistically significant, for a higher proportion of four-year-olds to make immature doll play choices than six-year-olds ($p.=.08$).

Table IV shows that the greater proportion of four-year-olds making immature choices was primarily accounted for by the bottle choice rather than the crib choice. A significantly higher proportion of four-year-olds than six-year-olds chose the bottle ($p=.01$). In this table a significantly higher proportion of four-year-old boys chose the bottle than four-year-old girls ($p=.05$).

In Table V, there are significant age-sex differences in parent doll choice ($p=.02$). A higher proportion of six-year-olds chose the same-sex parent doll than four-year-olds, i.e. a higher proportion of six-year-old boys than four-year-old boys chose the father doll, and a higher proportion of six-year-old girls than four-year-old girls chose the mother doll.

An inspection of Table V indicates that the six-year-olds chose the parent doll more discriminately than the four-year-olds. A smaller proportion of six-year-olds than four-year-olds chose the mother for Situation 10, Construction, choosing the father instead, and a larger proportion chose the mother for Situation 15, Play. The six-year-olds seem to have decided that it is the function of fathers not mothers to help one build things.

Thus, as would be predicted on the basis of the usual pattern of child de-

TABLE I—Age and Sex Differences for Dependency Scores

Source	df	MS	F
Sex	1	.047	<1
Age	1	.426	6.45*
Sex x Age Interaction	1	.518	7.84*
Error	42	.066	

* Significant at 5% level of confidence.

TABLE II—Sum of Dependency Scores

	Age 4	Age 6	Total
Boys	10.15	3.84	13.99
Girls	9.62	7.77	17.39
Total	19.77	11.61	31.38

TABLE III—Percentage of Immature Choices (Crib over Bed, Bottle over Cup)

	No Immature Choice	One Immature Choice	Two Immature Choices	N
Boys—Age 4	8	33	58	12
Boys—Age 6	55	27	18	11
Girls—Age 4	36	29	36	14
Girls—Age 6	33	50	17	12

NOTE: Rounding errors prevented the sum of percentages from totaling 100 in all rows in Tables III and V.

TABLE IV—Frequency of Mature-Immature Choices

	Crib	Bed	Bottle	Cup	N
Boys — Ages 4	8	4	10	2	12
Boys — Age 6	5	6	2	9	11
Girls — Age 4	8	6	6	8	14
Girls — Age 6	8	4	2	10	12

TABLE V—Percentage of Mother Doll Choices in Situation 10 (Construction) and Situation 15 (Play)

	Neither	Situation 10	Situation 15	Both	N
Boys — Age 4	42	33	17	8	12
Boys — Age 6	64	0	36	0	11
Girls — Age 4	50	29	21	0	14
Girls — Age 6	8	17	58	17	12

velopment, the younger children gave more immature SDP responses than the older children. Perhaps it would not be so readily predicted that a higher proportion of the older than the younger children would choose the same-sex parent doll. However, if one conceives of parent doll choice as representing the parent toward which the child strives to identify, then these results seem reasonable. That is, as the child grows older he strives more strongly to identify with the same-sex parent. This interpretation of parent doll choice would help account for the fact that, in the father-absent study reviewed above, a higher proportion of father-absent than father-present boys chose the father doll more often than the mother doll (12). Perhaps the father-absent boy, virtually lacking a father, strives all the more to identify with the absent father, and consequently tends to choose the father rather than the mother doll.

CASE ILLUSTRATIONS

The following case illustrations are presented to give the reader some familiarity with the quality of SDP responses, some notion of the kinds of material the clinician using the SDP has to work with, and an impression of the nature of the relationship sometimes found between SDP responses and data obtained from other sources. In order to cover a wide variety of illustrative responses, brief excerpts are taken from a number of SDP protocols rather than treating one record in depth. The resulting over-simplification is not intended to imply that this instrument should be used in a "cook book" fashion. Neither are these illustrations presented as interpretive guides, for that is beyond the scope of this particular paper.

The *Ss* used in these case illustrations are drawn from two sources. One group of *Ss* were administered the SDP as part of an evaluation at a university medical center child guidance clinic (henceforth referred to as Child G *Ss*). They were referred because of psychological disturbances. The second group were administered the SDP as part of a research project at a university medical center pediatrics outpatient clinic (henceforth referred to as Ped OP *Ss*). No *S* was included in the Ped OP group who came to this university medical center pediatrics out-patient clinic because of a psychological disturbance, but only those who came because of physical illness. Most of the illustrative material coming from sources other than the SDP, used for comparison with the SDP responses, was obtained from a semi-structured interview conducted with the mothers. In some of the Child G cases, material was also obtained from the social worker's interview with the parents and from the psychiatric interview with the child.

The first three examples concern various mature-immature choice patterns (cup - bottle, bed - crib, toilet - potty). In actual clinical use one would never consider the choices in isolation as is done here for simplicity's sake.

The first *S* for consideration is Child G subject *A*. Subject *A* is an eight-year-11-months old boy displaying infantile speech and immature behavior, such as crying easily. *A* considers boys too rough and plays mainly with girls. He is quite close to the mother and relies heavily on her for support, guidance, and protection.

A, despite his age, made all immature choices on the SDP, viz. crib, bottle, and potty.

Ped OP subject *B* is a four-year-eight-months-old girl referred because of recurring upper-respiratory infections. The mother did not feel that *B* feared the doctor, and reported that *B*, on her own, made an effort to be especially clean in preparation for the physical examination, whenever she was to see the doctor. The mother reported that *B* is content to play alone, but, on the other hand, adjusts well to groups also. She referred to *B* as an "easy going child."

This seemingly mature girl, although much younger than *A*, made all mature choices, viz, bed, cup, and toilet.

Child G subject *C* is an encopretic seven-year-11-months-old boy who, according to the mother, "started having trouble with his BMs just before he started going to school."

C made two mature choices (bed and cup). However, the one immature choice that this encopretic boy made was the choice of potty instead of toilet.

The next three examples illustrate patterns of parent doll choices. In the study with the SDP, which was treated in greater detail above, it was found that a significantly higher proportion of father-absent than father-present Norwegian children chose the father doll more often than the mother doll (12). In this study the absent fathers were Norwegian sailors whose occupation kept them away from home for long periods of time. Clinical experience suggests the hypothesis that high father doll choice accompanies father-absence in other than the specific conditions surrounding Norwegian sailor families. High father doll choice seems particularly likely where father-absence is coupled with maternal rejection.

Ped OP subject *D* is a five-year-11-months-old girl whose epileptic seizures are well controlled. She is an illegitimate child who has never lived with her father. The mother admits hostile feelings toward this unwanted child. When *D* was asked, "What do you think makes children sick?" she replied, "Because they don't like their mothers."

D chose the father doll in seven out of eight possible choice situations.[5] This proportion of father-doll choices is very rare, particularly with girls.

Ped OP subject *E* is a 10-year-three-months-old hemophiliac boy who has been separated from his real father since age two due to divorce. A stepfather has been in the home since the boy was six years old. The mother's strong feelings about *E* in connection with his illness are characterized by her comments, "It's all my fault. I'm the carrier. I shouldn't have had him in the first place. I've often had the idea, well, maybe he would be better off dead."

E chose the father doll seven out of eight possible times. In Situation 9, Sleeping Arrangement, *E* placed the boy and the father in the double bed, and the mother doll alone in the single bed.

Child G subject *F* is a nine-year-10-months-old boy who was referred for infantile speech. The father, in this case, is in the home. The mother reported that the child demands a great deal of attention from her. He, in turn, brings the mother candy. Whenever the mother kisses his 15-year-old brother, *F* asks her whether she loves him as much as she does his brother.

F chose the mother doll in all eight choice situations.

Ped OP subject *G* is used to illustrate peer choice in Situation 2A, Peer Choice. *G* is a six-year-11-months

[5] There are seven SDP situations in which the *S* is presented with the choice of mother or father doll. Parent choice is also scored in Situation 9, Sleeping Arrangement, making 8 possible choices in all. In this situation the parent doll placed in the same bed with the ego-doll, is considered the chosen parent. In addition, when the ego-doll is placed alone in the single bed, the parent doll placed on the side of the double bed next to the ego-doll is considered the chosen parent.

old boy referred because of an upper-respiratory infection.

G's mother reported that he likes to dress up in dresses and go to his girl friend's house. "He could play by the hour with her."

G chose the girl doll in the Peer Choice situation.

Sometimes a specific situation will illuminate a specific problem. Child G *H* is an eight-year-nine-months-old girl who was referred because of encopresis. The mother reports that "she don't seem to know even after her bowels have moved."

H gave the following response to Situation 8A, Toilet:

"Mommy takes the little girl to the potty and she can't go. She doesn't need to, but Mommy makes her go anyhow."

Typically, as the next case illustrates, a particular dynamic pattern may be reflected repeatedly throughout the test protocol, and in a variety of forms.

Ped OP *I* is an 11-year-old epileptic boy whose mother is blatantly overprotective. She so infantilizes *I* that the father complains that she "packs *I* around on a silver platter." Her protective feelings are reflected in her relating an incident in which *I* feared his hair would be cut in preparation for an EEG examination. The mother related that she reassured *I* that they would not cut his hair. The mother exclaimed to the interviewer, "No, my gracious, they couldn't mess that curl up. That would be tragic."

This 11-year-old boy reflected overprotection in Situation 1A, Crib or Bed Choice, by responding as follows:

I chose the crib and said, "He looks like a very young boy." He then made the mother choice and said, "Mother usually comes in first and checks the child. She has it say a prayer and kisses it goodnight."

In 8A, the Toilet Situation, *I* responded as follows:

I chose the potty. He said, "Mommy would unstrap his (the boy doll's) pants and he'd have to go. She would strap his pants again and wash his hands for him."

Situation 9, Sleeping Arrangement, was as follows:

The mother doll, rather than the father doll, was placed in the double bed on the side closest to the boy doll in the single bed. Then *I* said, "He (the boy doll) wants to go to the rest room or wants water. 'Mother, mother, I want water.' She goes and gets it for him."

In another situation the boy doll was said to cry, and in another to ask the mother doll whether she brought him a toy from town.

J is a four-year-five-months-old Child G boy brought to the clinic by the paternal grandmother because of infantile speech. The grandmother said that she has taken care of *J* since his birth because his mother (her daughter-in-law) neglected him, e.g., "She wouldn't even give him his bottle." The grandmother describes the mother's "temper tantrums" during which she has become so angry that she tears buttons off clothes. *J* was described as being so afraid of seeing his mother during her infrequent visits that he would bite his fingernails when he saw her approaching.

J's SDP responses show both oral preoccupation and fear of the maternal figure. In Situation 7B, Separation, *J* gave the following response:

"Boy hide. Mommy come back. She eat that boy. He all gone. She eat daddy. He all gone."

Fearful oral preoccupation was reflected in two additional situations. In Situation 9, Sleeping Arrangement, he said, "Bogy man eats them up," and in Situation 10B, Night Fears, *J* responded, "Bogy man ate the little boy up."

Subject *K* is a nine-year-six-months-old Ped OP girl who was referred because of vulval vaginitis. Four years previously, when she was five, a baby brother was born to whom she showed extreme hostile rivalry. The mother reported, "She (subject *K*) said I couldn't bring him (the baby) home

from the hospital. When we got home we couldn't keep her in the same room with him. She would bite him or pinch or slap him all the time if she got the chance. She'd try to push him off my lap. If she got scissors she would cut the nipple off his bottle."

The mother reported that *K* and her brother now get along "wonderful."

However, in SDP Situation 5B, Reprimand, *K* said:

"Well, she comes home and fights with her baby brother. Daddy says, 'Don't fight with him.' He makes her go to bed. He (brother) is only one-year-old. She lays there and cries, and father tells her not to cry, and they all go to bed. She started crying for her mother, and mother says, 'Be quiet,' and she says, 'No.' The girl hit her father and mother. Mother spanked her, and she shoved her brother out, and her father and mother spanked her again."

Situation 6, Parental Relationship, is structured so that the parent dolls are alone. After S relates what transpires while the parents are alone, the child doll is introduced. At this point *K* said:

"The girl tells her father she got all A's today. Then she tells mother, 'I got all A's today.' Mother and father say, 'O.K.' They ask the girl where her brother is. The brother tells mother and father he got all A's today. The girl asks her mother if she likes her, and mother says 'no' so she'll cry."

In Situation 7A, Separation, *K* said, "She goes to the neighbors and the little boy asks, 'Where's your brother?' She starts fighting."

K, in this situation, chose the mother as the parent doll she wished to return to the ego-doll first. *K* said:

"The mother wonders where the little brother is, and the girl says, 'I don't know.' The girl says, 'Give me a drink.' Mother gives her a drink and she throws the glass on the floor and goes to bed."

SUMMARY

This paper described the Structured Doll Play test (SDP), which is designed specifically as a projective doll play technique for use with children. In this paper, literature was reviewed which suggests that a structured approach may elicit more clinically useful material than free doll play, and that the SDP has been developed in keeping with doll play research findings.

Research findings from a study of father-absence with the SDP which suggest concurrent and construct validity were reviewed. An investigation of the SDP responses of 49 four- and six-year-old middle-class children was presented.

Concurrent validity in the SDP is suggested in this study by the discrimination between groups (age and sex comparisons).

A significantly higher proportion of four-year-olds than six-year-olds received high dependency scores. A significant age-sex interaction showed six-year-old boys particularly low on dependency. Four-year-olds tended to make more immature doll play choices than six-year-olds. A significantly higher proportion of four-year-old boys than four-year-old girls chose the bottle. There were significent age-sex differences in parent doll choice, with a higher proportion of six-year-olds than four-year-olds choosing the same-sex parent doll, and choosing the parent doll more discriminately.

Construct validity is suggested by the logical relationship between these findings and developmental patterns.

Research and clinical results suggest that the SDP may well meet the requirements of a useful projective test with children.

REFERENCES

1. Ammons, R. B., & Ammons, H. S. Parent preferences in young children's doll-play interviews. *J. abnorm. soc. Psychol.*, 1949, *44*, 490-505.

2. Driscoll, G. P. *The Driscoll Play Kit*. New York: The Psychological Corporation.

3. Duess, Louisa. Étude experimentale des phénoménes de résistence en psychoanalyse infantale. *Ztschr. f. Kinderpsychiat.*, 1944, *11*, 1-11.

4. Graham, E. E., & Whitmore, Lillian E. *The use of dramatic play for diagnosis and therapy*. Denver, Colorado: Whitmore Drama Kits, 1954.

5. Lillie, H. E., & Lillie, Marcella B. An experimental study of the difference between subjects' responses in a free and controlled play situation. Unpublished master's thesis, Univer. of Denver, 1951.

6. Lynn, D. B. An investigation of hypotheses basic to a concept of relative intensity of interaction as applied to Structural Doll Play test responses. Unpublished doctor's dissertation, Purdue Univer., 1955.

7. Lynn, D. B. Development and validation of a Structural Doll Play test for children. *The quart. Bull., Indiana Univer. Medical Center*, Jan., 1955.

8. Lynn, D. B. Structured Doll Play test manual. Unpublished manuscript (mimeographed), Indiana Univer. Medical School Library, Indianapolis, Indiana, 1957.

9. Mosse, Hilda L. The Duess Test. *Amer. J. Psychother.*, 1954, *8*, 251-264.

10. Phillips, Ruth. Doll play as a function of the realism of the materials and the length of the experimental session. *Child. Develpm.*, 1945, *16*, 123-143.

11. Robinson, Elizabeth F. Doll play as a function of the doll family constellation. *Child Develpm.*, 1945, *17*, 99-119.

12. Tiller, P. O. Father absence and personality development of children in sailor families. *Nordisk Psykologi's Monogr.*, 1958, Ser. No. 9.

Received September 16, 1958.
Revision received December 18, 1958.

Relationship between Story Length and Situational Validity of the TAT[1]

CAROLINE TAYLOR MACBRAYER
Mooresville, N. C.

Significant changes in Thematic Apperception Test (TAT) phantasy have been reported following experimentally produced increases in aggression (3), depression (5, pp. 454-455), sexual arousal (4), hunger (2), achievement motivation (15, 16), and affiliation motivation (1, 19). In scoring TAT stories to reveal these changes, which theoretically result from projection of experimentally aroused motivation into phantasy, various scoring methods have been used. Bellak (3, 5, pp. 445-457) simply counted the number of aggressive words in TAT stories under conditions of no criticism and under conditions of criticism of Ss' stories by E, and also counted the number of depressed words and phrases under post-hypnotically induced feelings of depression, as compared to normal. No account was taken of length of stories, data being in terms of absolute numbers of aggressive words or depressed words and phrases, rather than in terms of proportions of total words.

Milam (17) has demonstrated that male college students tell significantly longer TAT stories when the examiner is "as demanding, rejecting, and threatening to the subject as possible" (17, p. 222), or "as accepting, courteous, and friendly as possible" (17, p. 222), as compared to neutral conditions in which the examiner is efficient, routine, and noncommital, mean numbers of words per story being 162.80, 159.53, and 112.48, respectively (17, p. 224). Milam's results strongly suggest that increased motivation results in increased length of TAT stories, and with increased story length, the frequency of almost any type of word or thema might well be expected to increase. In order to demonstrate the situational validity of the TAT, therefore, it is necessary to show that an increase in the specific type of phantasy experimentally induced occurs independently of increased story length.

McClelland et al. (15, 16) have attempted to control the factor of story length in their work on arousal of the achievement motive by limiting TAT stories (written) to five minutes each. Average story length under these conditions was found to be about 90 words (15, p. 404), and these authors feel that " . . . content of the stories is not systematically related to their length" (16, p. 212). Relevant data presented are mean words per story for 28 achievement-oriented high school subjects ($M = 77.2$) as compared to 18 neutral similar high school subjects ($M = 68.3$), the difference between the means not reaching statistical significance at the 5% level (16, p. 212). Similar time limits were placed on TAT stories in experimental arousal of hunger (2), the affiliation motive (1, 19), and sexual motivation (4). Variation in story length is certainly reduced by a five-minute time limit, but whether this is an adequate control of story length, or even a desirable one if longer stories are actually more revealing of needs and desires[2], is questionable. Projection of experimentally aroused achievement (15) and affiliation motives (1, 19) was measured by frequency counts of items which increased significantly under arousal conditions. In Clark's (4) scoring of TAT stories for manifest sex imagery

[1] This research was conducted at Davidson College.

[2] Length of TAT stories which are unrestricted by a time limit has been used in some studies (10, 14) as a criterion of excellence of performance.

after experimentally induced sexual arousal, sexual activity was scored just once for each story. In scoring, priority was given to that type of imagery in each story which was biologically most sexual, the three types being: "(a) primary—explicit or implicit evidence for sexual intercourse, (b) secondary—evidence for the occurrence of such secondary sex activity as kissing, dancing, or fondling, (c) tertiary—characters in the stories perceived as sweethearts, on a date, courting, in love, and so on, but not engaged in either primary or secondary sexual activity" (4, p. 47). Symbolic sexual imagery, on the other hand, was scored by Clark (4) for each story as "strongly" present (weight of 2), "some" symbolism present (weight of 1), or none (weight of 0), depending on to what extent themas which could be interpreted as sexually symbolic were present. Clark's scoring thus gives frequency of sexual imagery more weight in the symbolic than in the manifest sex score, variations in story length being thereby relatively more important in the symbolic score.

The present experiment was designed to determine whether criticism results in a significant increase in length of TAT stories, and if so, whether increase in aggressive phantasy, as measured by numbers of aggressive words in the stories, is chiefly a function of increased story length, rather than being due to projection of increased aggression which has been experimentally induced by criticism. For purposes of this experiment the assumption was made that criticism does arouse aggression, but no assumption was made concerning the nature of the motivational variable or variables intervening between criticism and increase in TAT story length. It was assumed that if increase in aggressive words with criticism is primarily a function of increased length of TAT stories with criticism, four other types of phantasy, presumably unrelated to the experimentally induced motivation of aggression, would show similar increases and that frequency of each of the five types of phantasy would correlate significantly with story length. The specific hypotheses to be tested are as follows:

1. Criticism significantly increases the length of TAT stories.

2. While criticism increases the absolute number of aggressive words in TAT stories, it does not increase the proportion of aggressive words relative to the total number of words in TAT stories.

3. Increase in number of aggressive words in TAT stories with criticism is due to increase in length of TAT stories with criticism, rather than to the projection of experimentally aroused aggression into phantasy.

Method

Subjects were 20 psychologically naïve college freshmen, 10 males[3] and 10 females. It was assumed that Ss would more freely project socially disapproved feelings of hostility and aggression into TAT phantasy in the presence of student Es than a faculty E (18), so two male upperclassmen served as Es[4] throughout the experiment. Only the two Es and the particular S being tested were in the room at any time during the experiment. The procedure was similar to Bellak's (3), each S telling six TAT stories, the first three under normal or no-criticism conditions, and the last three under conditions of criticism by the Es. The TAT pictures used were: 18GF, 18BM, 13MF, 3GF, 10, and 3BM, this order being used with one-half of the Ss of each sex, and the reverse order with the other half. Instructions were as follows:

This is a test of imagination, one form of

[3] Data for one male S are not included under Results (N=19 instead of 20), because length of his stories under the first condition, without criticism, fell above the upper limit of the range for numbers of words per story reported by Dana (6) for psychotics.

[4] The author wishes to express appreciation to L. M. Mayfield and E. I. Stevens, who served as Es.

intelligence. I am going to show you some pictures, one at a time, and your task will be to make up as dramatic a story as you can for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking, and then give the outcome. Speak your thoughts as they come to your mind. No other person will know who has told what story, so do not be afraid to speak openly and frankly. Do you understand?

The same E made some encouraging remark to each S after the S completed one practice story in response to TAT picture 7GF before the experiment began, and also after S's first and second stories during the experiment. These remarks were "good," "fine," "O.*K*!" (said approvingly) or *"That's* the idea!" After the third, fourth and fifth stories, criticism patterned after that of Bellak (3, p. 361) was made of S's stories as follows:

[After the third story], "Your stories are not as good as they could be. They are illogical and unstructured. You lack imagination and ideas in your stories. Try and see if you can do better on this one." [After the fourth story], "I hope you do not mind if I am frank, but your stories still are not very good. In fact, I would go so far as to say that they are about the worst that we have had, so let's see you improve on this one!" [After the fifth story], "We have one more picture here for you, though I'm afraid you are almost hopeless. Anyway, put everything you have into this one!"

As soon as each S had completed his sixth story, he was told that he had been fed standard comment, was made to feel that he had done well, and was thanked for his cooperation. Ss spoke into a microphone and were aware that their stories were being recorded on tape. No names were used and full anonymity of results was assured each S by the Es.

Aggressive words and depressed words and phrases in all TAT stories were listed independently by two judges. There was a high percentage of agreement between judges, and where disagreements occurred, a mutual decision was reached after discussion. Total number of aggressive words and total number of depressed words and phrases were thus obtained for each S's three TAT stories under criticism and no-criticism conditions, as well as total number of words in each S's three stories under each condition. Two judges also scored all stories for number of manifest sex words (4), number of symbolic sex themas (4), and number of optimistic words and phrases, the mean for the two judges being used as the score. Coefficients of correlation between scores by the two judges ranged from .86 to .94. Optimistic words and phrases were defined for purposes of this experiment as those falling within one or more of the following categories:

1. *Hope*: things will, or probably will get better or work out.
2. *Faith*: prayer, faithful wife, help from above, etc.
3. *Kindness*: consideration for others or for pets; sympathy, support given to others.
4. *Patience*: only when stated or implied that patience is a virtue.
5. *Happiness*: includes all kinds of happiness except illicit sex *if* clearly stated or implied that the illicit sex experience makes person feel guilty, unhappy, or anxious.

In addition to numerical scores which were obtained for each S under the two conditions, as outlined above, each numerical score was also expressed as a proportion by dividing the numerical score by the total number of words in S's three stories from which the numerical score was obtained.

RESULTS AND DISCUSSION

Results for male and female Ss are combined, since the Sign-Rank Test of differences showed no significant sex differences in scores except a significantly larger number of manifest sex words for males. Table I reveals that story length increased with criticism, the difference between the means of total numbers of words with and with-

TABLE I—Mean Numerical and Proportional Scores for Three TAT Stories with and without Criticism

(N = 19)

Type of Score	Numerical score M without criticism	Numerical score M with criticism	r	t	Score as a proportion of total number of words p without criticism	Score as a proportion of total number of words p with criticism	t
Total words	280	432	.86**	3.90**			
Aggressive words	9.2	15.1	.07	2.25*	.033	.035	.26
Depressed words	15.5	23.0	.73**	2.78*	.055	.053	.38
Manifest sex	6.0	10.7	.50*	2.47*	.021	.025	.89
Symbolic sex	9.3	17.2	.68**	5.61**	.033	.040	1.45
Optimistic words	8.3	17.6	.61**	3.78**	.030	.041	2.00

* Significant at better than .05 level.
** Significant at better than .01 level.

out criticism being significant at better than the 1% level. This confirms the first hypothesis, that criticism significantly increases length of TAT stories.

The number of aggressive words also increased significantly with criticism, while the proportion of aggressive words did not (see Table I), which confirms the second hypothesis. At the same time, each of the four other numerical scores, namely, depressed words, manifest sex words, symbolic sex themas, and optimistic words, also increased significantly with criticism, while none of the proportional scores did so. From these results one might conclude that increase in aggressive words with criticism, as well as increase in the four other numerical scores, is a function of increased length of TAT stories with criticism. Further analysis of the data, however, reveals that this is not true and that hypothesis 3 is incorrect.

Table II shows that the only score which does not correlate significantly with story length is number of aggressive words under conditions of criticism. Further proof that increase in aggressive words with criticism is not a function of increased story length, as increases in the other four scores apparently are, is seen in Table I. Here the correlation between scores with and without criticism is significant in every case, including total numbers of words, except in the case of aggressive words, in which case the correlation is .07. Both lines of evidence, therefore, reveal the same picture and substantiate the conclusion that, while increases in depressed words, manifest sex words, symbolic sex themas, and optimistic words in the present experiment are a function of increased length of TAT stories with criticism, increase in number of aggressive words is actually due to projection of increased aggression which was experimentally produced by criticism.

TABLE II—Correlations between Total Words and Five Numerical Scores for Three TAT Stories with and without Criticism

(N = 19)

Type of score	Without criticism	With criticism
Aggressive words	.66**	.38
Depressed words	.68**	.84**
Manifest sex	.64**	.66**
Symbolic sex	.79**	.92**
Optimistic words	.67**	.92**

*Significant at better than .05 level.
**Significant at better than .01 level.

Results of the present experiment, then, confirm the situational validity of the TAT, despite the fact that the proportion of aggressive words in relation to total number of words was not significantly increased by experimentally induced aggression. This does not, of course, solve the more pressing problem of general TAT validity, which concerns sensitivity of the TAT to enduring personal characteristics. Lindzey and Herman (13) feel, however, that " . . . it is reassuring to find that the test [TAT] is demonstrably sensitive to situational or temporary

fluctuations, for this increases the plausibility of our expectation that the test is sensitive to similar enduring characteristics" (13, p. 39).

Another conclusion which seems justified on the basis of present results is that length of TAT stories is often an important factor in their interpretation. This is clearly seen in the results of the present experiment, where criticism is followed by a significant increase, not only in aggressive phantasy, but also in depressed, sexual (both manifest and symbolic), and optimistic phantasy in TAT stories. Most striking is the fact that the increase in optimistic words is actually more significant than the increase in aggressive words (see Table I). If the important factor of story length were ignored, it is easy to see how the results of the present experiment might be ludicrously interpreted as evidence that criticism induces happiness or optimism which is projected into TAT phantasy. When any type of words, themas, or other units (6, 7, 8) is used in scoring, length of stories becomes a vital factor because the longer the story, the better its chances of receiving a high score. When rating scales are used (9, 11, 17, 18), frequency as well as intensity of affect (or sometimes frequency as an index of intensity of affect) is ordinarily used in assigning ratings to TAT stories. With this type of scoring, too, the longer the story, the better its chances of receiving a high rating. Caution is therefore urged in applying any method of scoring to unusually short or unusually long TAT stories. Particular care is needed in working with abnormal populations, especially when their TAT performances are to be compared with those of normals, since Dana (6) found that story length for five TAT stories significantly differentiated a normal ($M = 595.8$) from a clinical population of neurotics ($M = 431.6$) and psychotics ($M = 420.1$). Range of numbers of words also showed a systematic trend, increasing from 350-1025 for normals, to 225-1195 for neurotics and to 165-1445 for psychotics (6). This raises the question of whether, in a study such as that of Gluck (9), wide variations in length of stories by 30 psychiatric Ss could account for the lack of relationship found between amount of hostility in six TAT stories and degree of overt hostility or aggression under stress.

It may be that a relatively short time limit, such as that used in measuring n Achievement (15, 16) and n Affiliation (1, 19), with TAT stories written, will be a sufficient control of length for group experimental work with normal adult Ss (12). McClelland et al. (16) feel that they have succeeded in adequately dividing their male college population only into high and low achievement groups, or at most, into high, middle, and low achievement groups (15, p. 406). In individual predictive work, however, especially diagnostic clinical work, and in work with other than normal adult populations (children, psychotics, neurotics, mentally deficient, etc.) short time limits are obviously inappropriate because they offer an additional handicap to the already difficult task of securing productive phantasy. It may be concluded that in scoring TAT stories, some method of taking into account the length of stories is desirable, and under certain circumstances, is a necessity.

SUMMARY

An experiment was performed to determine whether increase in number of aggressive words in TAT stories following experimentally produced increased aggression through criticism is due to increased story length. It was found that story length did increase significantly with criticism, as did five other scores: numbers of aggressive, depressed, manifest sex, and optimistic words and symbolic sex themas. When these five scores were expressed as proportions of total words, however, there were no significant increases from no-criticism to criticism conditions. All scores correlated significantly with story length except number of aggres-

sive words under conditions of criticism. Each score under no-criticism conditions correlated significantly with the same score under criticism conditions except number of aggressive words. It was concluded that, while increases in the other four scores were a function of increased story length with criticism, increase in number of aggressive words was due to projection of experimentally induced aggression, thus substantiating the situational validity of the TAT.

References

1. Atkinson, J. W., Heyns, R. W., & Veroff, J. The effects of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, *49*, 405-410.
2. Atkinson, J. W., & McClelland, D. C. The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, *38*, 643-658.
3. Bellak, L. The concept of projection, an experimental investigation and study of the concept. *Psychiat.*, 1944, *7*, 353-370.
4. Clark, R. A. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 44-57.
5. Crafts, L. W., Schneirla, R. C., Robinson, Elsa E., & Gilbert, R. W. *Recent experiments in psychology*. (2nd ed.) New York: McGraw-Hill, 1950.
6. Dana, R. H. Clinical diagnosis and objective TAT scoring. *J. abnorm. soc. Psychol.*, 1955, *50*, 19-24.
7. Dana, R. H. Cross validation of objective TAT scoring. *J. consult. Psychol.*, 1956, *20*, 33-36.
8. Fine, R. Manual for scoring scheme for verbal projective techniques (TAT, MAPS stories and the like.) *J. proj. Tech.*, 1955, *19*, 310-316.
9. Gluck, M. R. The relationship between hostility in the TAT and behavioral hostility. *J. proj. Tech.*, 1955, *19*, 21-26.
10. Light, B. H. Comparative study of a series of TAT and CAT cards. *J. clin. Psychol.*, 1954, *10*, 179-181.
11. Lindner, H. Sexual responsiveness to perceptual tests in a group of sexual offenders. *J. Pers.*, 1953, *21*, 364-374.
12. Lindzey, G., & Heinemann, Shirley H. Thematic Apperception Test: individual and group administration. *J. Pers.*, 1955, *24*, 34-55.
13. Lindzey, G., & Herman, P. S. Thematic Apperception Test: a note on reliability and situational validity. *J. proj. Tech.*, 1955, *19*, 36-42.
14. Lubin, N. M. The effect of color in the TAT on productions of mentally retarded subjects. *Amer. J. ment. Defic.*, 1955, *60*, 366-370.
15. McClelland, D. C. Measuring motivation in phantasy: the achievement motive. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955, Pp. 401-413.
16. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
17. Milam, J. R. Examiner influences on TAT stories. *J. proj. Tech.*, 1954, *18*, 221-226.
18. Mussen, P. H., & Scodel, A. The effects of sexual stimulation under varying conditions on TAT sexual responsiveness. brief report. *J. consult. Psychol.*, 1955, *19*, 90.
19. Shipley, T. E., Jr. & Veroff, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, *43*, 349-356.

Received November 25, 1958

The Blacky Pictures as Personality Measures for Undergraduate Areas of Specialization

MAX G. MAGNUSSEN
Veterans Administration Hospital, Cincinnati, Ohio
and University of Kentucky Northern Center, Covington, Kentucky

The stimulus for this research was a study by Teevan (7) that concluded that correlations between personality and vocation discovered by Roe (5, 6) and Knapp and Goodrich (4) for professional groups can also be demonstrated during the period of undergraduate field of specialization. Teevan used the Blacky Pictures as his diagnostic instrument and administered the test in a group setting. He mentioned in the article that there were two deviations from the standard procedure: "(a) up to 15 seconds longer than the standard two minutes were allowed on the spontaneous stories; (b) the inquiry questions, rather than being flashed on the screen one at a time with a time limit for each question, were printed at the bottom of each page. Ss were instructed not to glance at them until told to. An observational check and questioning at the end of the session showed good observance of these instructions." He failed to specifically report the deviations in clinical procedure of: (1) projecting the test stimuli on a screen, and (2) the examination being administered in a group.

The purpose of Teevan's experiment was to determine whether or not personality factors are correlated with choice of a major field in college. Three broad groupings of literature, social science, and science majors were compared on scores derived from the Blacky Pictures and several significant differences were found between these groups. The results were interpreted to demonstrate that correlations between personality and vocational groups can be found during the period preceding entrance into a profession.

This study attempted to utilize the Blacky Pictures as the personality measure in a more orthodox manner of administration and to determine if the results perviously reported by Teevan could be reproduced.

PROCEDURE

The Blacky Pictures (2) were individually administered to sixty male students from the University of Kentucky (Lexington campus and Northern Kentucky Extension Center). The subjects (Ss) ranged in age from 19 to 23 years. The major fields were divided into the same divisions used by Teevan, with twenty Ss from each of the academic areas of specialization. Division 1 (categorized as literature) was comprised of students (N=20) from the English department. Division 2 (classified as social sciences) contained students (N=20) from majors in government, economics, and history; and Division 3 (labeled science specialty) was composed of students (N=20) majoring in biology, geology, chemistry, and physics. In an effort to obtain Ss that were seriously involved in their educational field, only students who had achieved a B plus grade average in the area of concentration were used.

RESULTS AND DISCUSSION

The results of the present investigation are summarized in Tables I and II. The findings are presented in a manner that will permit direct comparison with the reported results of Teevan (7). Table I presents the significance of the differences between the three divisions when Blacky Picture disturbance scores are correlated with choice of major undergraduate specialty.

Table II shows the percentages of the students in the three divisions that gave disturbed and neutral responses

TABLE I—The Relationship Between Blacky Picture Disturbance Scores and Choice of Major Field

Blacky Picture Category	p	Ranking of Division by degree of Blacky Disturbance Scores		
		Most	Middle	Least
Oral Eroticism	.05	1	2	3
Oral Sadism	.05	2	1	3
Anal Expulsiveness	NS			
Anal Retentiveness	NS			
Oedipal Intensity	NS			
Masturbation Guilt	NS			
Castration Anxiety	NS			
Identification Process	NS			
Sibling Rivalry	NS			
Guilt Feelings	.06	2	1	3
Ego Ideal	NS			
Narcissistic Love Object	NS			
Anaclitic Love Object	NS			

Code: Division 1. — English; Division 2. — government, history, economics; Division 3. — biology, chemistry, physics, geology
NS=Not significant at .05 level.

TABLE II—Percentage of Each Division Showing Disturbed and Neutral Responses to Blacky Picture Dimensions

Blacky Picture Dimensions	Division		
	1 Literature	2 Social Sciences	3 Sciences
Oral Eroticism			
Disturbed	52	43	15
Neutral	48	57	85
Oral Sadism			
Disturbed	47	57	11
Neutral	53	43	89
Guilt Feelings			
Disturbed	48	58	28
Neutral	52	42	72

to the Blacky Picture dimensions which yielded statistically significant differences between the groups.

Table I reveals that the first division (English) was the highest of the three groups on Oral Eroticism. These findings corroborate Teevan's results, but with a lower degree (.05 as compared to .01 level) of statistical confidence. In addition, the findings are again consistent with Bergler's (1) observations of 36 people with literary abilities undergoing psychoanalysis and conclusions that writers are fixated at the oral erotic level. These results are also in agreement with Boyd's (3) findings of more disturbance at the oral erotic level in children that are good readers than in slow or retarded readers. Therefore, the current data from the Blacky Pictures are in accordance with the aforementioned investigator's categorization of literary persons as being essentially passive and dependent.

The second or social sciences division (government, history, economics) was highest on the other category that showed significant differences and the one category that closely approximated the .05 level of statistical significance. Oral sadism is purportedly related to aggression toward the mother; Guilt Feelings, as another dimension, is contingent on the introjection of parental standards. Teevan's "neat" constellation of a common element — emphasis on the mother, as reflected through Oral Sadism, Oedipal Intensity, and Anaclitic Love Objects was not found in its entirety. Only the Oral Sadism as aggression toward the mother (Blacky Picture #2) and the Guilt Feelings (.06

level of confidence) which might be consistently placed into the pattern are found to be significant differences.

Finally, Division 3, or the sciences (biology, chemistry, physics, geology) has the lowest disturbance score of the three divisions on most of the Blacky Picture categories. This finding is again consistent with Teevan's results and permits a conclusion that undergraduate males that migrate toward this division are less disturbed in psychosexual development than those who major in the other fields of specialization appraised in this experiment.

For those so inclined, the data present material for interesting clinical speculation and "hypothesizing" concerning personality development for various people in specific undergraduate areas of specialization. The results are fairly clear-cut and the present author is content to deal entirely with the obvious findings.

SUMMARY AND CONCLUSIONS

The present study attempted to determine whether the Blacky Pictures were adequate measures to appraise personality factors which had previously been found to be correlated with choice of undergraduate fields of specialization in college. The Blacky Pictures were individually administered to three categories of college majors consisting of twenty Ss in each grouping. Teevan's results (7) were replicated in respect to:

1. The literature division showing higher Blacky Picture disturbance scores on Oral Eroticism than the other two divisions.

2. The Social Science division revealing higher disturbance scores on Oral Sadism and nearly significant differences on Guilt Feelings. Teevan's findings regarding Oedipal Intensity and Anaclitic Love Object were not found to be significantly different in this particular sample of sixty University of Kentucky students. Teevan had the advantage of a larger N (eighty-five males) which might account for these differences in results.

3. The Science division exhibiting the lowest disturbance scores on practically all categories.

It is concluded that the Blacky Pictures are fairly adequate measures for appraising personality variables as these factors are related to undergraduate field of specialization. In addition, Teevan's results are generally reproducible and the current findings seem to lend further credence to the view that in early stages of academic specialization, particular individuals migrate to specific college majors because of personality factors within themselves. Teevan's group administration of the Blacky Pictures appears to afford benefits of a savings in time and loses little, if any, information which might be gained through individual administration of the test. This seems to be especially true if only the objective, scorable data are being considered.

REFERENCES

1. Bergler, E. *The writer and psychoanalysis.* New York: Doubleday, 1950.

2. Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, *39*, 3-99.

3. Boyd, R. D. Reading retardation as related to personality factors of children and their parents. Unpublished doctor's thesis, Univer. of Michigan, 1953.

4. Knapp, R. H., & Goodrich, H. *Origins of American scientists.* Chicago: Univer. Chicago Press, 1952.

5. Roe, Anne, A Rorschach study of a group of scientists and technicians. *J. consult. Psychol.*, 1946, *10*, 317-327.

6. Roe, Anne. Psychological examinations of eminent biologists. *J. consult. Psychol.*, 1950, *13*, 225-246.

7. Teevan, R. C. Personality correlates of undergraduate field of specialization. *J. consult. Psychol.*, 1954, *18*, 212-214.

Received November 24, 1958

Sexual Identification in Responses to Rorschach Card III

MARVEN O. NELSON, WILLIAM WOLFSON AND RALPH LOCASCIO
Middletown State Hospital, Middletown, N.Y.

A number of writers in the field of Rorschach research have suggested that responses to Card III of the Rorschach may be used to gain an understanding of the sex identification of the subject taking the test. (1, 2, 3, 4) The implication of these suggestions seems to be that the blot on Card III lends itself equally well to the perception of human forms of either male or female sex.

The present study is an attempt to test the hypothesis that Card III can be perceived equally well as either male or female human figures by subjects taking the test.

METHOD

Subjects: The data used in the study were obtained: (1) From the files of Rorschach protocols obtained from 159 males and 110 females registered at the Middletown State Homeopathic Hospital as regular patients or as referrals for psychiatric examination; and (2) from protocols obtained from 32 males and 52 females who were not hospital patients. This second group was mainly composed of persons who had applied for admission to the State Hospital School of Nursing, employees of the State Hospital, and a limited number of persons from nearby colleges and communities. Another group (3) was composed of 21 hospitalized and non-hospitalized persons who were known to have engaged in overt homosexual behavior. All protocols used in this study were selected from a larger collection of protocols on the basis of clear designation of human form as a response to Card III. Only those giving the human form response during the free association period of the test were used. The subjects from whom the protocols were obtained ranged in age from 10 years to 79 years, with 92 per cent of them between the ages of 18 and 50.

Procedure: The Rorschach protocols were obtained from the patient group as a part of their psychodiagnostic examination. A number of the non-patient group were given Rorschach tests as a part of the examination requirements for admission to Nursing School, and others were volunteers tested as a part of a research project.

Treatment of data: The subjects were divided into the following groups: Male non-patients (N=32), female non-patients (N=52), male patients (N=159), female patients (N=110), and male homosexuals (N=21). The chi square technique was used to test the significance of the distribution of responses within each group and to compare the various groups in regard to their differences in distribution of responses.

RESULTS AND DISCUSSION

The frequency distributions of responses are shown in Table I, along with the percentages as indicated. It should be noted that the percentage of those giving a "neutral" response (that is, those designating the human figure without indicating the sex) ranged from 28.57% in the case of the homosexual males to 50% in the case of the females. In all groups, the percentage of those designating "male" were greater than those designating "female". Among the homosexual males, however, the responses were almost equally divided between the "male" and "female" designations. The female patients showed the highest percentage of any of the groups in the preference for the designation of "male".

The chi-square analysis of the data

TABLE I. Human Figure Responses to Rorschach Card III Designated as "Male", "Female", or "Neutral"* by 347 Subjects.

Group	Responses "Male" N	"Male" %	"Female" N	"Female" %	"Neutral" N	"Neutral" %	Total N
Male Non-patients	12	37.5	6	18.7	14	43.8	32
Female Non-patients	19	36.5	8	15.4	25	48.1	52
Male Patients	61	38.4	31	19.5	67	42.1	159
Female Patients	47	42.72	8	7.3	55	50.0	110
Homosexual Males	8	38.1	7	33.33	6	28.57	21

*"Neutral" refers to those human figure responses where sex was not designated.

TABLE II. Chi Square Analysis of "Male" - "Female" Responses Showing the Significance of Distributions Within Each Group and Differences Between Groups.

Group	Compared with: M	F	MP	FP	HM
Males (M)	2.00				
Females (F)	.0679	4.16*			
Male Patients (MP)	.0004	.167	9.78**		
Female Patients (FP)	3.241	2.545	7.66**	26.981**	
Homosexual Patients (HM)	.465	.958	.656	4.500*	.62

* Significant at 5% level of confidence
**Significant at 1% level of confidence

TABLE III. Chi Square Analysis of Combined "Male-Female" Responses and "Neutral" Responses Showing the Significance of Distributions Within Each Group and the Differences Between the Groups.

Group	Compared with: M	F	MP	FP	HM
Males (M)	.50				
Females (F)	.129	.076			
Male Patients (MP)	.026	1.6016	3.954*		
Female Patients (FP)	.610	2.637	1.734	0.00	
Homosexual Males (HM)	1.341	2.233	1.366	3.013	3.841*

*Significant at the 5% level of confidence.

is shown in Tables II and III. Examination of the data in Table II reveals the following differences at the 5% level of confidence or above:

1. Female non-patients tended to designate "male" more often than "female".

2. Male patients tended to designate "male" more often than "female".

3. Female patients showed a stronger tendency to designate "male" than did any of the male groups.

Examination of the data in Table III shows the following differences at the 5% level of confidence or above:

1. More male patients designated the sex of the human forms seen on Card III than the number who gave "neutral" responses by leaving the figures with sex undesignated.

2. The homosexual male group gave sex designated responses more often than neutral responses.

CONCLUSION

The results of this study did not confirm the hypothesis that the Rorschach Card III blot can be perceived equally well as either male or female human forms. On the contrary, it suggested that, for all groups tested, the blot was more often seen as "male" or "neutral" than as "female". This was most evident in the female patient group.

The limited number of subjects in the non-patient male and the homo-

sexual male groups suggests further study of these groups would be desirable.

References

1. Due, F. O. and Wright, M. E. The use of content analysis in Rorschach interpretation: I. Differential characteristics of male homosexuals. *Rorschach Res. Exch.*, 1945, *9*, 169-177.

2. Klopfer, B.; Ainsworth, Mary D.; Klopfer, W. G. and Holt, R. *Developments in the Rorschach technique. Vol. 1.* Yonkers-on-Hudson, World Book Co., 1954. Pp. 332, 392.

3. Lindner, R. M. Analysis of Rorschach test by content. *J. Clin. Psychopathol.*, 1947, *8*, 707-719.

4. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *J. proj. Tech.*, 1949, *13*, 97-126.

Received October 16, 1958

The Experimental Induction of the Defense Mechanism of Projection[1]

ROBERT ROSENTHAL[2]
University of California at Los Angeles

Since Freud first defined projection as "the ego thrusting forth onto the external world whatever gives rise to pain internally" (12 v. 4, p. 78), the concept has generalized to such an extent that today almost any idiosyncratic perceptual behavior runs the risk of being labelled projection. Bellak probably perceived this when he worriedly agreed with Sears that projection was the most inadequately defined term in all psychoanalytic theory (3). Actually it would seem that a good many writers define projection rather carefully and adequately (1, 3, 7, 15, 30, 31, 32, 35), but many of us generate confusion by our varying usages of the term. The process whereby the concept of projection has picked up all its surplus meaning is discussed in more detail elsewhere (24).

Considering the importance of the concept, relatively little controlled research has been carried out on projection. Fabian (9), Rokeach (23), Sears (28), Zucker (36), and Norman and Ainsworth (20) defined projection operationally and except for Rokeach did find evidence for the operation of projection. Before accepting the conclusions of these investigators, it seems essential to review exactly what they did to see if we can be satisfied that it was indeed projection they were investigating. The basic designs of these studies were enough alike so that a general model may be constructed. A questionnaire purporting to measure various traits is filled out by a group of people all of whom know each other. Each person fills out the questionnaire once for himself and also once for every other person in the group according to how he feels that other person would fill in the questionnaire. Projection is said to occur when a person rates himself low on a trait on which others rate him as higher, and rates others higher on that trait than they rate themselves. Thus projection is defined by a discrepancy score. This is an operational definition and one cannot question the legitimacy of an operational definition —it is self-contained. One can, however, question the usefulness or appropriateness of any given operational definition. In these studies, because of the emphasis which the pathological features of projection have received, most of the traits investigated were undesirable ones. These undesirable traits may consciously or unconsciously have been denied by a large percentage of the subjects so that an empathic person correctly ascribing these traits to another would be likely to be accused of projection where he actually showed empathy. In this type of study everything hinges on knowing whether a person "really" has a given trait or not. Through ratings or through tests this is a difficult question to answer.

Perhaps a more important difficulty of these studies is that they in effect end by considering projection as a trait of which one can have more or less depending on the size of the discrepancy score. A trait is a *static* entity, but projection is widely thought to be a *process* (3, 12, 32). The definitions of projection referred to earlier imply this in no uncertain terms. Thus Freud speaks of "the ego thrusting," Bellak of "the ascription of feelings,"

[1] This study is based on a portion of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of California at Los Angeles. I wish to express my appreciation to my committee and especially to Drs. Bruno Klopfer and J. A. Gengerelli for their valuable advice and assistance.

[2] Now at the University of North Dakota. I want to thank Dr. Hermann F. Buegel for his reading of the manuscript.

and Symonds of "the reference of impulses." These are all active, process terms rather than static, trait terms. We may thus conclude that as far as these controlled observations are concerned the *process* of projection has been at best touched upon but by no means demonstrated to occur.

We may review briefly those investigations wherein an attempt was made to actually induce projection. Murray's classic study (19) showed that children who were frightened perceived photographs as more malicious than they did before their fear experience. While writers on the subject of projection have often quoted this study as evidence for the phenomenon, Murray himself considered this phenomenon to be "complementary apperceptive projection" rather than projection itself. Bellak (3) and Rodnick and Klebanoff (22) found that in response to criticism of story-telling and motor ability there was an increase in themes of aggression on the TAT. They cite this as evidence of projection insofar as the S's aggression was "projected" onto the TAT cards. Since Bellak defines projection as "the ascription of feelings and qualities of one's own to other people . . . " (3, p. 354) his finding might have been more direct and telling had he found that criticized Ss perceived more criticized figures on the TAT. Nevertheless, these studies are the ones which yield the most suggestive evidence to date for the existence of the process of projection.

In scrutinizing carefully the literature on projection it would seem most appropriate to demonstrate the process of projection by inducing it. If we can induce it, we not only have evidence for its occurrence, but we are also assured that it will remain a process, in our thinking, instead of being viewed as a static trait-like entity.

It might be well here to restate Symonds' definition of projection as the "reference of impulses, thoughts, feelings, and wishes originating in the person himself to persons and objects in the outside world" (32, p. 296). For experimental purposes it is necessary to limit this definition somewhat since impulses, feelings, wishes, and thoughts would be too awkward to investigate simultaneously. Evidence for the occurrence of the process of projection will be considered to have been advanced if known feelings can be induced in persons who can be made to show that they refer these experimentally induced feelings to others.

Sears (29, p. 235) has said: "Perhaps not the most compelling, but certainly the most pervasive of all socially determined rewards are those leading to feelings of success or failure." Since it was felt that this would be especially true for college freshmen, it was decided to choose the feelings of success and feelings of failure as those to be induced in the experimental Ss.

Method

Subjects

One hundred and eight subjects were divided into three groups. There were thirty-six male paranoid-diagnosed patients from a neuropsychiatric hospital, thirty-six male students and thirty-six female students all enrolled in first year psychology classes. The ordinary objection to highly selective sampling due to using volunteers was not operative since the patients were requested by their physicians to participate and the college Ss were obliged to serve time as research Ss as part of their course requirements (5). The mean age of the patient group was 36.0 years and their mean Shipley-Hartford I.Q. 110.3 (118.3 when only vocabulary I.Q. was used). The mean age of college men was 20.1 and of college women 18.4. Thus while the three groups were roughly equated for I.Q. the patient group ran about seventeen years older than the mean of the college Ss. The paranoid group, primarily paranoid-schizophrenics were selected by their physicians on the following criteria: 1) that they be of at least average intelligence, 2) that they be in sufficient touch with the world to be able to understand moderately dif-

ficult instructions, and 3) that they be clearly and typically paranoid. All the physicians contributing Ss included "projection" as one of their chief criteria of a paranoid disturbance. Paranoids were included since so much of the literature links this specific disorder with the defense mechanism of projection carried to pathological extremes (18, 25, 31). In addition, if any group of psychotic patients, especially schizophrenics, are to be included in a study with college students, it seems most appropriate to use paranoid patients since these tend to be more nearly of the same order of intellectual functioning. In general, too, paranoids are more like normals than are the other schizophrenic nosologies (16, 17, 21).

College women comprised a separate group in view of Schilder's (27) analysis of the different meaning of success and failure to men and women in our culture. In addition, Zucker's work suggests a sex difference in the projecting process (36).

Materials

The Photo Card Sets. The stimulus materials upon which the Ss were to project their feelings were photographs of men's and women's faces cut from advertisements in a weekly news magazine. They ranged in size from 30 x 46 mm. to about 42 x 68 mm. and were mounted on 3 x 5 inch white cards in such a way that most Ss thought they were photos taken solely for the purposes of this study. Twenty-eight of the most ambiguous photo cards were assigned to one of two sets of fourteen each. The sets were "matched" by the method of random assignment (8). Each subject rated one set of photo cards before and one set after the experimental treatment thus serving as his own control.

The Rating Scale. On a rating scale running from —10 to +10 subjects were asked to rate the face on each photo card according to the degree of success or failure they felt was expressed by that face. A —10 rating was made when the S felt the face was of a person who had just been informed of a very severe failure in some important unspecified task. A +10 rating was made when the S felt the face was of a person who had just been informed of a very extreme success in some important unspecified task. A —1 rating was made when the S felt the face was of a person who had just been informed of a very mild failure. A +1 rating was made when the S felt the face was of a person who had just been informed of a very mild success. All other numbers occupied intermediate positions between —10 and —1 if negative and between +1 and +10 if positive.

The "Intelligence Test." In view of the paranoid's frequent preoccupation with intellectual matters and the college student's necessary preoccupation with these, it was felt that strong feelings might easily be aroused in all Ss in connection with performance on an intelligence test. Two pseudo-intelligence tests were constructed which resembled some of Wechsler's sub-tests (24, 34). One of the tests was analogous to Block Design while the other was very roughly analogous to Object Assembly.

Experimental Design

Experimental Treatment. The overall design of the experiment was to determine whether the S's experience of either success or failure on the "intelligence test" influences his ratings of success and failure on the photo cards. Each of the three groups of subjects; the patients, the college men, and the college women, were further subdivided into three experimental treatment groups: a success group, a failure group and a neutral (control) group. The failure group was included because it is widely felt that unpleasant feelings are projected. It is thus the primary experimental group which on the basis of the literature should show the projection process most clearly. The success group was included because of a few writers' (28, 32, 33) assertions that there is a "projection of

desirable and good" feelings and impulses onto other people as well as a "projection of the bad" (32, p. 297).

The success and failure groups were read the same initial instructions (24). They were told that the purpose of the study was to see if there was any relationship between "empathy" and "intelligence." Insofar as could be determined no subject suspected the true intent of the experiment. The success and failure subjects were then shown how to make their ratings of the photo cards and told that they would afterwards take an intelligence test based on "the best intelligence test we use in psychology today." Instructions to the neutral group include no mention of an intelligence test. Throughout the four items which comprised the "intelligence test" the "success" S was allowed to finish and was warmly complimented. The "failure" S, however, was not permitted to finish any item and the E's remarks became increasingly solicitous, e.g., "Perhaps we should try an easier one."

After having undergone the success or failure treatment, the S took another "empathy" test wherein he rated the other matched set of fourteen photo cards. If the hypothesis of projection were correct we would expect the failure group to see more failure on the second rating than on the first. In terms of the ratings assigned each photo card this would mean that the S assigned a lower algebraic number to the photo presented after failure. We may call this change in rating a shift score. The shift score is positive in sign if more success is seen after treatment and negative in sign if more failuse is seen after treatment. If the hypothesis of the "projection of the good" were correct we would expect the success group to see more success on the second rating than on the first. The neutral S, instead of taking an "intelligence test," was permitted to read the school paper between the first and second ratings. Presumably the neutral group's second ratings should not differ significantly from their first ratings. Before leaving this section it must be pointed out that in order to avoid any harmful psychological effects of the failure Ss without giving away the intent of the experiment, an extension of the experimental design became necessary. All failure Ss were given the "success" treatment after their second ratings and told that only their best performance would "count." Thus, while many failure Ss were quite upset by the failure treatment, none of them remained disturbed after their subsequent success experience.

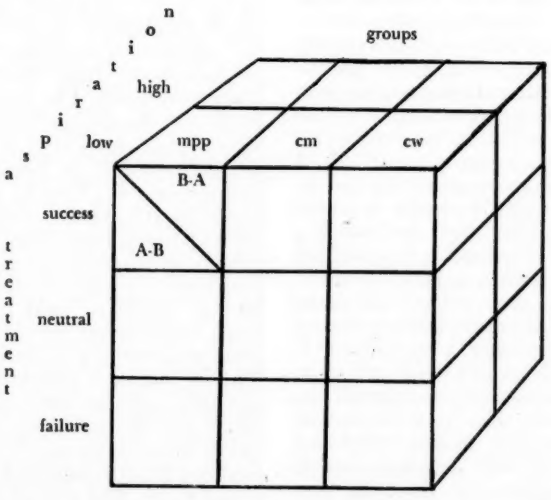
Level of Aspiration. Although it was felt that success and failure on an intelligence test would represent an emotionally "real" experience for the Ss, the possibility still existed that some or all of the Ss would remain uninvolved (13, 14). To this end, all Ss were given the Cassell Group Level of Aspiration Test (6). Frank and others (2, 6, 10, 11, 26) point out that ego involvement is related to a high level of aspiration and that this level of aspiration is a fairly global characteristic of a person. Thus a high aspirer would be expected to have a greater negative reaction to the experience of failure. In terms of the experiment, then, even if we found no projection by any of the treatment groups, we might nevertheless find that those Ss who had a high level of aspiration did show projection. We would then be in a position to state that although projection was not generally induced, it *was* induced in those cases where the S was ego-involved and cared about his success or failure. This in itself would be considered appropriate evidence for the occurrence of projection. Accordingly, each of the three groups of subjects: the paranoid patients, the college men and the college women were divided into high aspiring Ss and low aspiring Ss. In each group the eighteen Ss scoring highest on the level of aspiration measure became the "high aspirers" and the eighteen low scorers became the "low aspirers."

The Factorial Design. To summarize, three groups of Ss were each sub-

divided into three experimental treatment groups and each of the resulting nine groups were then divided into high and low aspiration levels. This 3 x 3 x 2 factorial design is illustrated in Figure 1. Since it was not possible to be certain whether the two photo sets of the "empathy" test were really identical, half of the subjects in each of the eighteen cubicles rated photo card set A first while the other half rated photo card set B first. This counterbalancing of photo card set presentation was designed to cancel out any differences due solely to the difference in photo card sets A and B. This of course expanded the factorial design to a 3 x 3 x 2 x 2. It will be remembered that the data going into each cubicle are shift scores. For example, a before-treatment rating of +2 and an after-treatment rating of +6 yields a shift score of +4. A before-treatment rating of −6 and an after-treatment rating of −2 also yields a shift score of +4.

RESULTS

The analysis of variance based on the shift scores from the first to the second "empathy" test indicates whether any changes have occurred which might be attributed to treatment, to group (or nosology) and to the two levels of aspiration. The eighteen cubicles formed by the three groups, three treatments and two levels of aspiration yield variance estimates based on three main effects, three simple interactions and one triple interaction. Since each cubicle was divided into two semi-cubicles differing in order of photo card set presentation, the within cubicle variance was broken down into within semi-cubicle variance. There were four within cubicle sources of variance: (a) "persons within semi-cubicles," was the variance due to individual differences; (b) "card order within semi-cubicles" was the variance due to the fact that the fourteen cards of each photo set



Key

mpp – male paranoid patients
cm – college men
cw – college women
A-B – photo card set order
B-A – photo card set order

FIGURE 1 — Diagram of 3x3x2x2 Factorial Design

TABLE I—Analysis of Variance Based on Pre-Posttreatment Shift Scores

Source of Variance	*df*	Mean Square	Significant *Fs*
Aspiration	1	0.76	
Treatment	2	3.78	
Groups	2	1.01	
Asp. x Treat.	2	32.26	
Asp. x Groups	2	7.43	
Treat. x Groups	4	16.00	
Asp. x Treat. x Groups	4	38.18	
Persons within semi-cubicles	72	31.01	
Card order within semi-cubicles	468	46.48	1.60*
Interaction card order x person	936	29.17	(Error term)
Residual	18	197.97	
Total	1511		

* Significant at the .01 level

differed from each other and from the fourteen cards in the other photo set; (c) "interaction between persons and card order within a semi-cubicle" was the variance due to the differing cards and card sets' differential effects on individuals. Since this source of variance was the variance left after the known sources of within cubicle variance were removed (persons within semi-cubicles and card order within semi-cubicles), it was used as the error term.[3] (d) A residual variance representing the interaction between the order of photo card presentation and the differing semi-cubicles remained as an additional source of within cubicle variance. Table I shows the results of the analysis. None of the among cubicle sources of variance showed a significant F. Only the "card order within semi-cubicles" source of variance was significant. Our conclusion thus must be that by the experimental design employed projection was not demonstrated to occur. In addition it must be concluded that the difference between the two photo sets in amount of shift was markedly significant.

[3] For working out many of the difficulties in this rather complex statistical treatment I am extremely grateful to Prof. J. A. Gengerelli.

DISCUSSION

In this type of personality research, there is always the possibility that the S's knowledge that he is participating in an experiment makes it difficult to elicit responses (e.g. projection) which are typically thought to occur in "real" situations posing a threat to the ego. Thus the simple knowledge of his being in an experiment may be enough to serve as a kind of antidote to experimentally induced stresses. This conceivable failure to ego-involve the Ss may account for the experimental failure to demonstrate the occurrence of the projection process. In addition to this more general possibility there appear to be three others more specific to this experiment which might account for this negative finding.

The first of these is that the experiment may have in some way been inappropriate to the problem at hand. Thus it may have been unrealistic at this stage of our knowledge in this area to attempt a precision of technique which may have been inconsistent with the precision of our backlog of empirical information from which hypotheses must be derived. The second possibility is that the definition of projection must remain, for a time at least, a "sensitizing concept" as Blumer has used this notion (4). That is, the clinician knows what he means when he uses the term and feels reasonably sure that his communicants hold roughly the same sensitization to the concept. Under this view it would be held that until we learn more about projection there is little use in trying to define it further. This is indeed a possibility but, it is felt here, a remote one. It appears more likely that by increasing the precision of our concepts we are put into a better position to gain empirical information about them. The first of the possibilities accounting for this failure to induce the projection process is a question of experiment while the second is a question of conceptual clarification. The possibility that these two objections may not account for the negative

findings leads us to the possibility that the process of projection may not occur as different writers have formally defined this concept. It may be then, that the reason for this concept's having gathered so much surplus meaning is chiefly that we just do not find the simple, uncluttered, clearly defined process in operation. Thus, in addition to our continuing clinical observations of the process of projection it would seem worth while to continue experimental attacks upon the problem in the hopes that we can clarify the process and perhaps specify the conditions under which it is likely to occur.

SUMMARY

The purpose of this experiment was to test the hypothesis of projection, operationally defined, by a direct attempt to induce the process. Accordingly, 108 Ss composed of 36 hospitalized paranoid patients, 36 college men, and 36 college women were further divided into three sub-groups receiving success, failure, or neutral experience on an "intelligence test." The nine subgroups thus formed were further divided into high and low aspiring Ss within each sub-group, thus yielding an 18 cubicle factorial design. Data were based on shift or difference scores between an S's pretreatment ratings of photos as more or less successful or more or less unsuccessful, and posttreatment ratings. It was postulated that if the hypothesis of projection as defined by Bellak, Symonds, and others were tenable, the success group should see (rate) more success after treatment; and the failure group should see (rate) more failure after treatment when both groups are compared to the neutral (control) group. Paranoid patients were included among the Ss since the literature suggests that they are more prone to project. College women were included because of data suggesting sex differences in projection. High aspirers were hypothesized to become more ego-involved with an intelligence test experience and thus perhaps show the process of projection more clearly. The analysis of variance showed none of the among cubicle sources of variance to be significant. Thus there was no support for the hypothesis of projection as defined by Bellak, Symonds and others within the data of this experiment. Possible explanations for this finding were discussed.

REFERENCES

1. Allport, G. W. *Personality, a psychological interpretation.* New York: Holt, 1938.
2. Bayton, J. A., & Whyte, Esther. Personality dynamics during success-failure sequences. *J. abnorm. soc. Psychol.*, 1950, *45*, 197-200.
3. Bellak, L. The concept of projection, an experimental investigation and study of the concept. *Psychiat.*, 1944, *7*, 353-370.
4. Blumer, H. *Sensitizing concepts.* Paper read at Amer. Sociol. Soc., Berkeley, Aug., 1953.
5. Brower, D. The role of incentive in psychological research. *J. gen. Psychol.*, 1948, *39*, 145-147.
6. Cassell, R. N. *Manual: the Cassell group level of aspiration test.* Beverly Hills: Western Psychological Services, 1952.
7. Coleman, J. C. *Abnormal psychology and modern life.* Chicago: Scott, Foresman & Co., 1950.
8. Edwards, A. L. *Statistical analysis for students in psychology and education.* New York: Rinehart, 1946.
9. Fabian, W. A., Jr. An investigation of the relationship between measures of insight and measures of projection and distortions in ratings. *Dissert. Abstr.*, 1954, *14*, Part I, 711. (Abstract)
10. Frank, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, *47*, 119-128.
11. Frank, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, *47*, 285-293.
12. Freud, S. *Collected Papers.* London: Hogarth, 1953, Vols. 3, 4.
13. Hilgard, E. R. Success in relation to level of aspiration. *Sch. & Soc.*, 1942, *55*, 423-428.
14. Hoppe, F. Erfolg & Miserfolg. *Psychol. Forsch.*, 1930, *14*, 1-62. (under) Lewin, K. *Untersuchungen zur Handlungs-und Affektpsychologie, IX.*
15. Ichheiser, G. Projection and the motebeam mechanism. *J. abnorm. soc. Psychol.*, 1947, *42*, 131-133.
16. Miller, D. R. Pathological reactions to threat of failure. *Amer. Psychologist*, 1949, *4*, 274. (Abstract)
17. Miller, D. R. Reponses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, *46*, 378-387.
18. Murray, H. A. *Explorations in personality.* New York: Oxford Press, 1938, Ch. 4.
19. Murray, H. A. The effect of fear upon

estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, *4*, 310-329.

20. Norman, R. D. & Ainsworth, Patricia. The relationships among projection, empathy, reality and adjustment, operationally defined. *J. consult. Psychol.*, 1954, *18*, 53-58.

21. Rickers-Ovsiankina, Maria. Studies of the personality structure of schizophrenic individuals: I. the accessibility of schizophrenics to environmental influences. *J. gen. Psychol.*, 1937, *16*, 153-178.

22. Rodnick, E. H. & Klebanoff, S. G. Projective reactions to induced frustration as a measure of social adjustment. *Psychol. Bull.*, 1942, *39*, 489.

23. Rokeach, M. Studies in beauty, II. some determiners of the perception of beauty in women. *J. soc. Psychol.*, 1945, *22*, 155-169.

24. Rosenthal, R. An attempt at the experimental induction of the defense mechanism of projection. Unpublished doctoral dissertation, Univ. of California at Los Angeles, 1956.

25. Rosenzweig, S. An outline of frustration theory. In Hunt, J. M. (Ed.) *Personality and the behavior disorders*. New York: Ronald Press, 1944, Vol. I, Ch. 11, 379-388.

26. Rotter, J. B. Level of aspiration as a method of studying personality, III: group validity studies. *Charact. & Pers.*, 1943, *11*, 255-274.

27. Schilder, P. Success and failure. *Psychoanal. Rev.*, 1942, *29*, 353-372.

28. Sears, R. R. Experimental studies of projection: I. attribution of traits. *J. soc. Psychol.*, 1936, *7*, 151-163.

29. Sears, R. R. Success and failure: a study of motility. In McNemar and Merrill (Ed.) *Studies in personality*. New York: McGraw-Hill, 1942, 235-258.

30. Sears, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Council Bull.*, 1943, No. 51.

31. Slotkin, J. S. Culture and psychopathology. *J. abnorm. soc. Psychol.*, 1955, *51*, 269-275.

32. Symonds, P. M. *The dynamics of human adjustment*. New York: Appleton, 1946.

33. Van Lennep, D. J. The four picture test. In Anderson, H. H. & Anderson, Gladys. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.

34. Wechsler, D. *The measurement of adult intelligence*. Baltimore, Williams & Wilkins, 1944.

35. Weiss, E. Projection, extrajection, and objectivation. *Psychoanal. Quart.*, 1947, *16*, 357-377.

36. Zucker, K. B. An experimental investigation of correlates of projection. Unpublished masters dissertation, Western Reserve Univ., 1952.

Received December 7, 1958

The Integration of Determinants and Content in Rorschach Interpretation[1]

DAVID SHAPIRO
Austen Riggs Center

1.

First of all, I would like to state my understanding of the question that is presented to us at this symposium on "*The Relative Importance of Determinants and Content in Rorschach Interpretation.*" I assume that by "importance," simple empirical validity is *not* meant. That would be a perfectly legitimate question—by which method of interpretation one could obtain the best predictability in a given situation. But I am not sure how much would be learned from that question, just because I am convinced that no really decisive answer could be obtained, that the answer would vary depending on the variable to be predicted and also on the skill and the direction of interest of the predictors. If the question were to be simply concerned with matters of everyday, practical importance, I think it would be easier to argue for the greater importance of determinants, and for a very simple reason: the Rorschach happens to be an instrument the design of which is directed toward indicating certain formal elements of individual perceptual tendencies rather than elucidating verbal content. This is its unique contribution and its historical significance, and although it is theoretically possible that an instrument designed for one purpose would later turn out to perform, with optimal efficiency, another purpose, this would be essentially an unlikely coincidence.

However, I assume that the question is a different one. Namely, is there a difference in the importance of these two factors which is intrinsic to them and to the Rorschach test; are there theoretical reasons why, ultimately even if not at present, one approach is likely to reveal decisively more of what we are interested in than the other? This is the question that, it seems to me, must be answered negatively. I hope that my position will not be taken for fence-straddling on this matter, because my argument essentially is for a different point of view, namely that each of these—determinant and content interpretation — is absolutely indispensible to optimal understanding of the other, and that the question of "relative importance" therefore cannot be answered one way or another for good reasons. What I will present to you could hardly be called a rationale for this position. It is only a fragmentary sketch of the reasons for a point of view which, however, I have found absolutely necessary in clinical work.

2.

What we obtain in a Rorschach response or a Rorschach protocol are essentially final, external products of internal processes. It is in the nature of such final external products, and especially those that reflect, as the Rorschach protocol does, highly complex internal processes, that they have in one particular sense a certain ambiguity. I mean by this simply that there is no one to one correspondence between a given final product and a specific inner process. This is not to question the determinism of such final products, such as an expressed thought or a piece of behavior, but only to say that the product itself does not necessarily reflect the processes that preceded it in any unique way. While I think this is

[1] Originally presented at the symposium on "The Relative Importance of Determinants and Content in Rorschach Interpretation" sponsored by the New York Society for Projective Techniques, May 17, 1958.

one of the central problems of diagnostic work in general, it is also a commonplace, and it is very easy to illustrate:

Imagine that you walk along the street and hear a man singing to himself a popular, romantic love song. You listen carefully to his words; what do they tell you? As carefully as you may listen, the words will not tell you whether he is in love or isn't in love, whether he is happy or sad, whether he is rehearsing a tune for an audition he is about to have or whether he really means, literally, what he is saying. Any of these internal processes and many more might have resulted in this external product, and his words will not reveal which. In fact, if one would direct attention too closely to the content of the song, that very fact might well preclude any clear impression at all of the man's actual mood and interest.

Consider another example. Suppose one is waiting outside a telephone booth and notices that the man who is using the phone is talking in a very excited, rapid, almost hysterical manner. Perhaps the impression is of someone who is making a call which concerns some family emergency. But it might equally well be someone, with quite different feelings, trying hard to make a last minute plane reservation, or even someone telling a friend about a piece of very exciting and happy news. It is likely in this case that knowing the exact content of the man's speech would establish one or another of those possibilities as the actual fact. Such knowledge, of course, would then allow a fresh and more insightful examination of the man's excited manner.

It seems to me that this fact that a given, final external product and in particular a single dimension of such a product need not necessarily reveal in a unique way the internal processes that precede and cause it is the *general* fact that argues against weighing the importance of one dimension of the Rorschach against another. It seems that the Rorschach is such an excellent instrument *just because* it provides a number of dimensions, of which, incidentally, the determinants and the ideational content are only two of the major ones.

3.

I would like to make some points now concerning the nature of the Rorschach response and the processes behind it.

The final Rorschach image, the fully conscious and externalized or expressed Rorschach response is, in a way of speaking, not just the product of associative and perceptual operations or even of the interplay between them. That is, it is not merely a result. It is the terminal point of a process, but it is in itself, an integrative operation. The nature of, and the success or failure of, this integration is as we all know one critical factor in any interpretation. It is so clear that the integration of ideational content and the visual-perceptual factors is the task in the Rorschach and that the task is neither simply perceptual nor simply associative, that when this integration fails, for example, in failure of sufficient associative distance from or elaboration of the visual impression, as in color-naming, or in an incapacity to bind associative trends sufficiently to visual factors, we do not hesitate to consider this a reflection of serious pathology, or, in some cases, we may rule this out altogether as a response proper.

Nor do I believe it can be said, as I understand the so-called projective hypothesis to say, that the integration which constitutes the Rorschach task is only a *compromise* between associative trends and external visual requirements. Here again, I think that if in a Rorschach protocol one actually gets an impression of such a *compromise* between internal associative pressures and external, visual requirements, one marks it down as a less than adequate integration, as a response of less than good form-level, for example, and considers it to that extent a failure.

The visual qualities of the ink blots are not ordinarily a restriction on associative processes and a factor with which one compromises. The ink blot qualities are, within that situation, the vehicle of associative processes, their medium, and the perceptual organizing tendencies that are brought to bear on these visual qualities may be compared with a syntax without which these associative processes cannot be expressed. As with any expressive medium or syntax, this one — the visual qualities available together with the organizing tendencies brought into play—shapes and lends its stamp to the content which is expressed through it.[2] Thus, one cannot view Rorschach imagery as if, for example, it were dream imagery any more than one can view dream imagery as if it were waking daydream imagery, or that as if it were ordinary day to day discourse or conversation. It is in the nature of human equipment to see things, to organize visual experience, to coordinate it with memory, interest, knowledge and such, and the ideational content that appears in such circumstances cannot be most profitably viewed except as coordinated with that actual visual experience. Those who work intensively with the Rorschach appreciate this fact intuitively. It is more or less built into many aspects of our approach to the Rorschach protocol, and it is built into our recognition as pathological of any incapacity to satisfy the ordinary standards of such coordination between visual and ideational factors. Naturally, this is not to say that ideational content or mode of expression in the Rorschach situation is unrelated to ideation which appears in other contexts, that a great deal may not be learned about one from the study of another, or that certain principles or individual modes of organization may not be common to all.

The Rorschach, in addition to its ideational aspect, allows for a high degree of individual inclination or personal style in perceptual organization. We know of course that the individual perceptual style which we see in a Rorschach protcool, or rather than speak of a single perceptual style, the organization of a variety of styles and tendencies which may express themselves in a protocol, reflects broader modes of expression and of character organization. The recent experimental work in individual perceptual and cognitive styles, for example of Klein and his associates, is after all in its outlook not new to us. For us, *the formal characteristics of the Rorschach, including particularly the determinants, are the data with which we arrive at an understanding of individual perceptual style and organization.* The determinants have this meaning or none at all.

It could be argued with some justification that the determinants have not always clearly and explicitly had this meaning in the past. The various determinants have been attached, with a greater or lesser degree of empirical validity, to a hodgepodge of human traits which seem impossible to systematize. Color attached to affect as though the color itself actually contained some affect deposit and with insufficient understanding of the perceptual processes through which such a relationship may, or may not, be established; shading attached to anxiety; FM to impulses of some sort; M to empathy; and so on. In spite of this state of affairs, we still must assume that the determinants fundamentally have their meaning as reflecting formal aspects of perceptual style and organization, and that their further interpretive or indicator-meaning must ultimately be understood in that sense.

Just as content gets its sense and its tone through the medium and style of its expression, so the expressive style is unthinkable without content. It is, again, the whole sense and direction of visual and perceptual equipment to see *things,* and the nature of the things which are seen, the quality of the

[2] For a demonstration of this point, Werner and Kaplan's (1) work on expression in line drawings is valuable.

things that are of interest, in other words, the visual content, will give the perceptual style its final and definite shape. To use an analogy which is not too distant from our interest, one cannot be simply compulsive, or depressive, or repressive. Compulsiveness can exist only with a content of action or ideas, and the nature of these actions and ideas which means, also, the nature of the material available for them, will determine the final form of compulsive tendency.

4.

My aim is not to convince you that you should rely just equally on content and on determinant interpretation in Rorschach work. It seems to me that perfectly legitimate differences of interest, natural inclination or background will tend to cause each one to give more weight to one side than to the other. Also, there is no doubt that there are differences in this respect among different Rorschach protocols; some, quite legitimately I think, seem to call for more emphasis on content interpretation and others on determinant interpretation. The aim here is, rather, to show that there is no reason intrinsic to the test to consider one aspect more important than the other, and, in addition, to show that each aspect can be properly understood only in the light of the other.

I would like to illustrate this, not exhaustively, but with certain problems which I think are common or typical ones. I will take two sorts of problems for this purpose: 1) the problem of levels or the problem of reference in content interpretation; and 2) the problem of specificity in interpretation which relies on determinants primarily.

Let me give an example in order to clarify the meaning of this problem of levels or problem of reference in content interpretation. We are all familiar with certain Rorschachs in which there is a steady current of sado-masochistic imagery. This imagery, for instance, may include such themes as tearing, cutting, beating, kneeling, or other themes suggestive of sadistic action, humiliation, or the like. There are several interesting problems in connection with the interpretation of a content theme such as this, but for the time being we will consider only one. The action-mode suggested by this content, namely sadism, has a certain ambiguity, and it retains this ambiguity no matter how precisely, by tracing the content theme throughout the protocol, one is able to delineate its details. The ambiguity may be described in several ways. It may be described as an ambiguity as to subject and object —on which end of the sadistic action does the person who gives the response feel himself to be? Or one can put the same question in a different way and ask, is the action with which the response-giver is primarily identified a sadistic or a masochistic action?

It might be said that this supposed ambiguity is only apparent, and that a great deal of clinical experience as well as certain theoretical considerations make it very clear that the apparent ambiguity is only an expression of the fact that these two positions—subject and object, or these two interests, sadistic and masochistic—are always and without fail present together within an individual. Essentially this is true, but it only sharpens the problem and makes it more clearly a problem of levels, for the fact remains that there are individuals for whom the sadistic mode is by far closer to conscious attitude and behavior, others for whom the masochistic mode is closer to conscious attitude and behavior, and perhaps a third group who may in fact be characterized primarily by vacillation in attitude and behavior between the two positions depending on specific circumstances. The question is, how do we choose among these possibilities on the basis of Rorschach content.

The crux of this problem of levels, you will notice, is this: the imagery content, the fantasy or memory or interest elements that are utilized in the construction of the Rorschach image

or the response content are, or at least may be, exactly the same in all these cases. What one does not know, and what one may very well have, from the content dimension alone, a difficult time finding out, is the exact relationship of this imagery to the person who gave it, its sense and meaning to him, or the mood and attitudes from which it emerges. Thus, the meaning and the value of this content interpretation remains seriously limited, in the illustration, to speaking only of a general sado-masochistic mode, until one can answer these questions.

Where will the answer come from? It will come most often, I think, from an understanding of the formal qualities of perceptual organization that appear in the same Rorschach. The way in which the visual material is shaped and handled is another aspect or dimension of the responses whose imagery content one has been examining, and this other dimension will ordinarily give clues to the answers one needs, that is, to the prevailing attitudes and orientation in the context of which the imagery has its full meaning.

I cannot present here a case interpretation, but I would like to describe what I mean more specifically. The same content imagery—cutting, beating, etc.—will sound different and will mean something different in a Rorschach with a large number of responses, with a great predominance of pure form responses of excellent form level, with little or no color or only color that is decisively subordinated to form, different in such a record from what it will mean in a record of few responses, of proportionately fewer pure form responses and those of a gross, more careless quality, with a large number of color-dominated responses and the immediate and passive relationship to the stimulus that those tend to imply. What is being suggested here, of course, is an active, forceful, penetrating perceptual style, on the one hand, and a more passive, giving-one's-self-over to the visual material on the other. The same imagery content may appear in these two widely different contexts, but it will have different meaning in each.

This is only one of a large number of such problems of level. Phobic and counter-phobic issues contain somewhat similar problems. Broader dimensions such as masculinity-femininity and aggression-passivity will also need to be considered in the same light.

Take another type of problem that, again, for the sake of convenience, may be called a problem of levels. A 35-year-old man gave the following response to the usual area of card VII:

> "They look like two old, gossipy women. I noticed the dark shading (in their faces)—it looks like somebody has thrown blueberry pies in their faces, like in the old slapstick comedies."

It does not take much ingenuity to detect an aggressive aspect to this response. And in fact, this man's Rorschach contained a number of other responses similarly marked by aggressive imagery. We will assume now, which is quite justified in the circumstances, that these responses reflect an active and aggressive interest more than a subjective feeling of being passively aggressed upon. Assuming this, what do we learn from this sort of content and what do we *not* learn? It is again a fact of the type that I suggested in my analogies at the beginning that identical aggressive imagery, of this sort for instance, may be given on the one hand by individuals who would not hesitate a moment to *act* aggressively, physically, and on the other hand by individuals who would never seriously consider such action. Not that these latter individuals are necessarily sweet Pollyannas—if that were the case, they probably would not give much imagery of this sort in the first place. But there are individuals, for example those characterized by what is called isolation of affect, whose predominant mode is in fact an active and in a sense aggressive one, and who are quite capable of entertaining on a certain level a good deal of aggressive

imagery, but whose *attitude* toward that imagery is quite detached. This man, for example, was a lawyer, who was not tested as a patient and in whom there was no outstanding evidence of pathology. The response content was, as far as he was concerned, a little joke. The formal Rorschach picture, the determinants, in this case were quite typical of compulsive people; the form level was impeccable. It is true that this sort of imagery reflected a somewhat lofty and condescending, possibly teasing attitude toward women on the part of the subject. But that is quite a different story from raw aggression. One must conclude that content elements such as these which may in themselves be identical would mean very different things in this man's Rorschach and in another characterized, for instance, by one or two pure C's, unsteady form level, and such. The content elements would have a different setting, would be drawn from a different level, so to speak, in the organization of thoughts or fantasies or memories.

I am not at all sure that the problems which might be called problems of content reference are essentially very different from the problems I have just indicated. These problems of content reference revolve around several facts, the first of which is that many images, though not necessarily ambiguous or esoteric, may have such a variety of connotations that the interpreter, unless he is satisfied to be arbitrary, is mystified about their specific connotation on a given occasion.

Take the simple response "waiter" to card III. Is this an image with oral connotation, or with a connotation of servility and status, or one of formality and restraint, or some other? It is true that many times these questions can be answered by study of other content or of general content themes. But not always.

For example, there may be a general content theme around images of flowers and exotic decorations; does this have a connotation of artistic taste and sensitivity, or one of compulsive and precious pseudo-esthetic interest, or one simply of sensuousness? The formal qualities of the perceptual style may, again, give some clue as to the actual connotation for the subject or may at least limit the range of possibilities, and therefore contribute to the content interpretation which is most valid.

Or consider another example, this time a young man's Rorschach in which there was great emphasis on images of food, more so than I recall having ever seen elsewhere. One might think that the connotation of food is clear enough. Perhaps in a camp of hungry or half-starved people it would be. But in our milieu, though it *can have* connotations of oral or passive craving, it can also have connotations of stubborn negativism, or sensuous indulgence, or many others. This record, formally, was a poorly controlled one, rather wild and reckless in form and in color-use, but at the same time with many responses that bespoke a kind of narcissistic self-confidence through all the reckless and impulsive judgments. I would say that the connotation of food, to this subject, was much more one of indulgence, a spoiled-child's tyrannizing demandingness, than the depressive or lonely cravings it may sometimes suggest.

The point may be summarized by saying that the real meaning of content is largely dependent on the kind of thought organization from which it is drawn, the subjective attitude towards it, and the personal connotations which it may have. But the establishment of these facts, and therefore the most adequate content interpretation, will usually require an understanding of the quality of the perceptual organizing tendencies through which this content has gained expression.

5.

What about the use of content for determinant interpretation? I think that here it is necessary to distinguish

two types of problems, one involving practical problems which derive from the limitations of our present knowledge techniques, and the other involving problems more intrinsic to our material.

In speaking of the first type of problem, I have in mind the types and categories in which we classify or differentiate the formal qualities of perception, in other words, our scores. These simply are too gross for our needs, and the various traits which they are said to reflect, even if in that respect they are perfectly accurate, are likewise not nearly specific enough. High form level, low form level, pure color, form-dominated color or, control, affect, impulsiveness, and such—these dimensions, as significant as they are, are simply in themselves insufficiently specific. At least they are insufficient if one wants to go beyond more or less gross diagnostic evaluations in test work.

But there is a more important and perhaps more interesting issue involved, namely, is it in the nature of purely formal modes, such as the perceptual modes which are represented for our purposes here by the determinants, ever to have the degree and kind of specificity which we require in our work? Or to put the question another way, can we look forward to the day when purely formal interpretation will be sufficiently specific. My own belief is that we cannot. I have already suggested what seems to me the essential reason for this; that is, that, limiting ourselves for the present to speaking of formal modes of *perception,* these modes come into being or are manifest only with some sort of ideational content, and it is in fact this content which crystallizes their final and specific shape.

What does this mean concretely? We can find some illustrations in connection with the use of the color-determinant.

Consider three color responses, or, let us assume, pure color responses: blood, spilled (colored) ink, and urine stains. If these three responses are all scored pure C, then by definition within our present scheme of describing these formal qualities of perception, the three responses are not distinguishable from one another. It is not that I doubt that they are different; I am sure that they are different, not only in content but in the formal quality of the perception. But these are differences which may very well not be detectable, and in this particular case are certainly not detectable, on the basis of the formal score alone. Why should this be so? Essentially, I believe, the answer is not in an insufficiency of our scoring systems, but rather because in the final moment, when the perceptual tendency is integrated with ideational content, it receives its essentially unique quality and stamp which may, to one degree or another, be special to that moment and to that content. And the quality and tone of the content is the obvious clue to the final shape which the broad perceptual tendency has at that moment taken.

In my experience, the "blood" response is most typically found in hysterical people where it seems to have the connotation of high pressure of affect, the "spilled ink" response is found often in impulsive, narcissistic character disorders who, as far as affect in the ordinary sense is concerned, tend to be rather cool and even bland, and the "urine stain" in chronic schizophrenics where the same formal score and broad perceptual mode reflects a severe and pervasive blandness. I have very little doubt, myself, that the color at that moment of response *looks* different to subjects in each of these three categories. But the key to how it looks, to the *exact* sense and quality of the perceptual mode at that moment, will not necessarily come from the formal score. Nor would it solve the problem fundamentally if we had finer, more highly differentiated scores for these responses. That might be helpful, but no imaginable number of divisions could approach the number of *specific* qualities and tones which the broad per-

ceptual mode is capable of taking in final expression. Ultimately the best clue to how the visual stimuli are perceived at the moment of response will come from the specific content of the response.

Any number of other examples, and many striking ones, are easy to come by among color responses. How else, if not by the whole tone and sense of the content, could one get a clue to the vivid, compelling sensory quality of a response such as this one given by an acutely paranoid psychotic boy to the center red area of card III:

"Like bright headlights on a car—like when you're on a beach at night and a car points its headlights at you."

Or the different quality of vividness in a hypomanic woman's response to all of card IX: "An orgasm in color."

Determinants — major categories of formal perceptual qualities — tell us something in these cases. But the fact is, again, that it is simply not in the nature of these categories to delineate with sufficient sharpness the visual and perceptual operations which can be of value to us in understanding the specific condition of the subject.

Examples could, of course, be multiplied with other determinants. In connection with the form determinant, for example, we already include in our scoring one attempt to specify the perceptual tendency more sharply by reference to the adequacy of integration of visual and content factors. I am referring to form level scoring. And additional gains in sharper specification are made in scoring, for example, "vague form" separately. The advantages of such finer form scoring are parallel to the advantages of scoring "deterioration color" separately, and all such attempts necessarily consider aspects of content or integration of content and visual quality in their attempts to delineate the formal perceptual style more finely. These distinctions are always valuable, and there will be room for finer scoring differentiations of this sort up to the point where the scoring system will be too cumbersome. But at the same time, it will be well to recognize the fundamental quality of these problems and to recognize that adding scoring differentiations here and there will never enable us to obtain as adequate a view of the special perceptual tendency at the moment of response as will, often, the content of the specific image that has become the final and externalized shape of that tendency.

6.

I am well aware that I am not saying anything new to experienced Rorschach workers. I believe in fact that the method and approach that I have been describing is one of the operations implied in the term "clinical intuition" as applied to Rorschach interpretation. But it is of value to make such operations explicit as far as we can. For one thing, they can throw additional light on such a question as the one we are discussing today; perhaps secondly, they may lay to rest some of the experimental research methods which tally up contents or determinants as though each and every time these had the same subjective meaning; and finally, they may suggest a more rigorous clinical application of this approach. I think that it may be just because the content dimension used in the way I have described is sometimes overlooked, that formal, determinant interpretation tends to be too gross in its descriptions and tends to lack a sharpness that gives a sense of conviction that a real human being is described. On the other hand, I expect that it is inadequate use of the formal, or determinant, dimension that causes content interpretation to be, in my opinion, almost always sharply detailed, but often incorrect.

I would like to make one final point. When we work actively with at least two dimensions of a production in this way, there is a special additional gain: we have a built-in criterion of adequacy. From this point of view, any Rorschach interpretation is less than adequate if it does not succeed in an

integration of every formal and every content aspect, and as long as *one* is not clear in the light of the *other,* we know that there is something missing. I would not say that this approach guarantees correctness, but I think it makes correctness much more likely than any other method.

REFERENCE

1. Werner, H. and Kaplan, B. "Experimental Studies on Symbolization," Lecture at Austen Riggs Center, March, 1958.

Received December 16, 1958

ceptual mode is capable of taking in final expression. Ultimately the best clue to how the visual stimuli are perceived at the moment of response will come from the specific content of the response.

Any number of other examples, and many striking ones, are easy to come by among color responses. How else, if not by the whole tone and sense of the content, could one get a clue to the vivid, compelling sensory quality of a response such as this one given by an acutely paranoid psychotic boy to the center red area of card III:

"Like bright headlights on a car—like when you're on a beach at night and a car points its headlights at you."

Or the different quality of vividness in a hypomanic woman's response to all of card IX: "An orgasm in color."

Determinants — major categories of formal perceptual qualities — tell us something in these cases. But the fact is, again, that it is simply not in the nature of these categories to delineate with sufficient sharpness the visual and perceptual operations which can be of value to us in understanding the specific condition of the subject.

Examples could, of course, be multiplied with other determinants. In connection with the form determinant, for example, we already include in our scoring one attempt to specify the perceptual tendency more sharply by reference to the adequacy of integration of visual and content factors. I am referring to form level scoring. And additional gains in sharper specification are made in scoring, for example, "vague form" separately. The advantages of such finer form scoring are parallel to the advantages of scoring "deterioration color" separately, and all such attempts necessarily consider aspects of content or integration of content and visual quality in their attempts to delineate the formal perceptual style more finely. These distinctions are always valuable, and there will be room for finer scoring differentiations of this sort up to the point where the scoring system will be too cumbersome. But at the same time, it will be well to recognize the fundamental quality of these problems and to recognize that adding scoring differentiations here and there will never enable us to obtain as adequate a view of the special perceptual tendency at the moment of response as will, often, the content of the specific image that has become the final and externalized shape of that tendency.

6.

I am well aware that I am not saying anything new to experienced Rorschach workers. I believe in fact that the method and approach that I have been describing is one of the operations implied in the term "clinical intuition" as applied to Rorschach interpretation. But it is of value to make such operations explicit as far as we can. For one thing, they can throw additional light on such a question as the one we are discussing today; perhaps secondly, they may lay to rest some of the experimental research methods which tally up contents or determinants as though each and every time these had the same subjective meaning; and finally, they may suggest a more rigorous clinical application of this approach. I think that it may be just because the content dimension used in the way I have described is sometimes overlooked, that formal, determinant interpretation tends to be too gross in its descriptions and tends to lack a sharpness that gives a sense of conviction that a real human being is described. On the other hand, I expect that it is inadequate use of the formal, or determinant, dimension that causes content interpretation to be, in my opinion, almost always sharply detailed, but often incorrect.

I would like to make one final point. When we work actively with at least two dimensions of a production in this way, there is a special additional gain: we have a built-in criterion of adequacy. From this point of view, any Rorschach interpretation is less than adequate if it does not succeed in an

integration of every formal and every content aspect, and as long as *one* is not clear in the light of the *other*, we know that there is something missing. I would not say that this approach guarantees correctness, but I think it makes correctness much more likely than any other method.

REFERENCE

1. Werner, H. and Kaplan, B. "Experimental Studies on Symbolization," Lecture at Austen Riggs Center, March, 1958.

Received December 16, 1958

BOOK REVIEWS

Mira Y Lopez, Emilio, Myokinetic Psychodiagnosis, New York: Logos Press. 1958. (Translated from the French by Mrs. Jacques Dubois.)

This book, originally published in French in 1951 has been translated under the editorship of Leopold Bellak, Michael H. P. Finn, Leonard Small and Frances Bishop. It is a manual for the use of the Myokinetic Psychodiagnostic technique which is an unstructured graphomotor approach to the investigation of personality characteristics. Six sets of simple line figures are to be reproduced by the subject under varying conditions. Both hands are used for the drawings. For the main portions of his responses the subject is not permitted to see the movements he is making but must depend on kinesthesis. A special test booklet and a special table permitting controlled tilting are required.

The test is based on the now familiar concept that bodily movements reflect personality characteristics. The use of both hands provides an opportunity to make comparisons based on the belief that "The dominant hand is expected to show more ego expression in that it exercises more practical control; the weaker hand, lacking the same degree of voluntary corrective control, is looked upon for more instinctive unconscious expressions." There are brief comments about the interpretation of the graphic responses. These are based on linear and chronological measurement as well as some special pattern variations. Intelligence, depression and euphoria, intrapsychic coherence (integratedness) excitation and inhibition, anxiety and apathy, aggressiveness, introversion and extroversion are among the areas dealt with. There are also suggestions for use of the instrument for diagnostic differentiation. Some normative tables are provided based on populations other than those of the United States. Some statistical evidence of reliability and validity are included.

As the editors point out, in its present form the technique is too cumbersome for the practicing psychologist but certainly provides an intriguing opportunity for research. There has been a great deal of research outside this country but little here. Although there are other graphomotor approaches such as that by Gehl and Kutash, Rapkin, and others, Mira's organized approach, using both hands for the comparative purposes he does, is a distinctive one that merits further consideration by psychology in the United States. However, the instructions both for administration and interpretation are extremely meagre and at many points difficult to follow. At this time, as implied by the editors, the translation of the manual has its greatest value as a stimulant to those who would do research in the area of graphomotor expression.

MORTIMER M. MEYER
Los Angeles

W. J. Revers, *Der Thematische Apperzeptionstest (TAT).* Hans Huber; Bern and Stuttgart, 1958. 209 pp.

This is an elementary text on the TAT by a German psychologist. After the introduction, there is a description of the test, application and administration, including a supplementary interview and a discussion of various methods of diagnosis using the TAT. He concludes with a section on unresolved problems, and one on the psychocathartic side-effects (Nebenwirkung) of the TAT.

For American psychologists the book is of interest solely for comparative purposes. It contains no new material, and recommends some procedures, such as using the TAT to stimulate psychotherapeutic discussions, which most psychologists here would consider to be at least of questionable value.

Almost all the literature cited is from American sources. However, he has relied mainly on published books and relatively little on journal articles.

Inasmuch as a number of stories and protocols are cited from German sources, these could be used for comparison with stories from other countries by anyone interested in the "culture-bound" aspects of the TAT.

The book is Vol. 2 in the Encyclopedia of Psychology in Individual Presentations, a Hans Huber series. Vol. 1 is a translation of Gordon Allport's "Becoming."

REUBEN FINE

GENERAL NEWSLETTER

Alvarez Villar, A. & Narros Martin, G. La percepcion visual en las diversas entidades nosologicas psiquiatricas. *Revista de Psicologia general y aplicada,* 1958, *13,* 739-766.

A perceptual test, the Progressive Matrices and the Rorschach were administered to several diagnostic groups. Empirical findings and theoretical points are presented.

Schachter, M. Psicodiagnostico, con el test de Rorschach, del caracter oposicionista e inestable. *Archivos de Criminologia, Neuro-Psiquiatria y Disciplinas Conexas,* 1959, *7,* 14-21.

The author describes two Rorschach variables which he finds characteristic of the protocols of unstable and oppositional personalities.

Porteous, Hedy S. Enuresis: the symptom and the character structure. Its relationship to the oedipal conflict. *A Crianca Portuguesa,* 1958, *17,* 665-681.

The psychoanalytic theory of enuresis is documented with a discussion of drawings.

Dabrowski, K. La desintegration mentale comme facteur positif dans le developpement emotionel de l'enfant. *A Crainca Portuguesa,* 1958, *17,* 731-736.

Indications of positive features among signs of disintegration in children are discussed and related to the Rorschach.

Menano, Cecilia. Importance du travail colletif dans le dessin et dans la peinture libres. *A Crianca Portuguesa,* 1958, *17,* 809-810.

A brief report on the use of children's group participation in creating designs and paintings.

dos Santos, J. & Menano, Cecilia. L'acquisition de la forme dans les dessin et dans la peinture enfantines. *A Crianca Portuguesa,* 1958, *17,* 811-813.

A description of stages of development in children's free drawing and painting.

Diaz Arnal, Isabel. Valeur diagnostique du dessin des enfants deficients. *A Crainca Portuguesa,* 1958, *17,* 937-938.

Diagnostic clues in children's drawings.

Meili, R. Das Problem der Validation der Methoden der Personlichkeitsuntersuchingen. *Zeitschr. f. diagnost. Psychol.,* 1958, *4,* 283-303.

A theoretical discussion of validity studies and theor implications.

Lucena, J., Borges, J. C. C., Sette, P. E. & Souto, Maria do C. Le Bender Gestalt Test et quelques donnees d'analyses structurales chez des schizophrenes. *Zeitschr. f. diagnost. Psychol.,* 1958, *4,* 313-316.

Significant differences were found on the Bender between acute and chronic schizophrenics.

ANNOUNCEMENTS

TWENTIETH ANNUAL PROGRAM

of the

Society for Projective Techniques

CINCINNATI, OHIO

FRIDAY, SEPTEMBER 4, 1959

9:00 - 10:50 A.M.

Symposium: The Role of Stimulus Structure in Projective Techniques. Dr. Edwin S. Shneidman, Central Research Unit, VA Center, Los Angeles, Chairman.

Participants: Dr. Kenneth B. Little, National Institutes of Mental Health; Dr. Bernard I. Murstein, Univ. of Portland; Dr. Edith Weisskopf-Joelson, Purdue Univ.; Dr. Ira Friedman, Cleveland Psychiatric Institute.

Discussants: Dr. Jerome Kagan, Fels Research Institute; Dr. Joseph Zubin, N.Y. State Dept. of Mental Hygiene.

SATURDAY, SEPTEMBER 5, 1959

11:00 - 12:50 A.M.

Symposium: The Prediction of Overt Behavior Through the Use of Projective Techniques. Dr. Arthur C. Carr, New York Psychiatric Institute, Chairman.

Participants: Dr. William E. Henry, University of Chicago: "The TAT"; Dr. Bertram R. Forer, Los Angeles VA Mental Hygiene Clinic: "Sentence Completion Tests"; Dr. Max Hutt, University of Michigan: "Graphomotor Techniques (Bender-Gestalt Test & Draw-a-Person Test)"; Dr. Zygmunt A. Piotrowski, Jefferson Medical College: "Rorschach Test".

Discussant: Dr. Evelyn Hooker, U.C.L.A.

3:00 - 4:00 P.M.

Business Meeting, Society for Projective Techniques.

4:00 - 5:00 P.M.

Presidential Address, Society for Projective Techniques.

Dr. William E. Henry: Personality Development in Adulthood and Old Age.

5:00 - 6:30 P.M.

Dutch Treat Cocktail Party, Society for Projective Techniques.

APA CONVENTION INTERNATIONAL PROGRAMS

One of the highlights of the 1959 APA Convention in Cincinnati will be a Round Table on "Reciprocal Influences in International Psychology," scheduled for Saturday afternoon, September 5, from 2-4 p.m., in the Ballroom of the Hotel Sinton. Jointly sponsored by APA Divisions 1, 9, and 12, the Round Table will also have representation from the Interamerican Society of Psychology and the Society for Projective Techniques. Participants and their areas of discussion include Drs. Gordon W. Allport (Africa), Eugene L. Geier (Scandinavia), Roger Myers (Canada), Nevitt Sanford (Eastern Europe), and Emanuel Schwartz (Latin America), plus two visiting psychologists from abroad (still to be announced). Dr. Henry P. David is Chairman. Members and guests planning to attend the APA Convention will also be encouraged to participate from the floor.

Immediately following the Round Table, APA will again sponsor an Open Coffee Hour for Visitors From Other Lands in the Crystal Room of the Hotel Sinton. As before, the International Council of Women Psychologists has been asked to arrange hostess service.

MEMBERSHIP DECISIONS

The Society for Projective Techniques is pleased to announce the following membership decisions.

Fellows:

Farberow, Norman L.	Roe, Anne
Garfield, S. L.	Rosenzweig, Saul
Halpern, Florence	Sanford, R. Nevitt
Holtzman, Wayne H.	Shapiro, David
Horn, Daniel	Singer, Erwin
Kalinkowitz, B. N.	Snowden, Robert F.
Lazovik, David A.	Stein, Morris I.
Lindzey, Gardner	Tomkins, Silvan S.
Little, Kenneth B.	Treat, Wolcott C.
Molish, Herman B.	

Associates:

Akavia, Uriel	Jortner, Sidney
Berliner, Hildegard	Klein, Louis S.
David, Charlotte	Liutkus, Stankey
Ehrenreich, Gerald A.	Reens, Renee
Fabrikant, Benjamin	Schon, Martha
Greenberg, Nathan	Scott, Edward
Haworth, Mary R.	Tada, Harou
Hoshino, Akira	Tanaka, Fuzio
Isaacs, Mark	Zeev, Bracha

Student Affiliates:

Donoghue, John	Schoenborn, Elizabeth
Kleban, Morton	

BOOKS FOR REVIEW

The Executive Editor invites offers to re-

view the following books.

Bass, B. M. & Berg, I. A. (Eds.) *Objective approaches to personality assessment.*

Bellak, L., Ostow, M., Pumpian-Mindlin, E., Stanton, A. H., & Szasz, T. S. *Conceptual and methodological problems in psychoanalysis.*

Caligor, L. *A new approach to figure drawing. (With test materials for the Eight Card Redrawing Test.)*

Delay, J., Pichot, P., Lemperiere, T. & Perse, J. *The Rorschach and the epileptic personality.*

Harrower, Molly. *Personality change and development (As measured by the projective techniques.)*

Orr, Miriam. *Le test de Rorschach et l'imago maternelle.*

Szondi, L., Moser, U. & Webb, M. W. *The Szondi Test in diagnosis, prognosis and treatment.*

ERRATA

Three errors exist in the paper by Arthur B. Silverstein. Identification with same-sex and opposite-sex figures in thematic apperception. J. proj. Tech., 1959, *23*, 73-75.

All errors are in Table 1. The value of *t* for Male Succorance should be 1.59, not 1.61. The value of *r*, Opposite-Sex for Female Dominance should be +.32, not —.32. The value of *t* for Female Dominance (2.61) is significant at the .05 level, and should be starred.

If anyone wishes to dispose of back issues of the Journal, foreign psychologists who lack the funds to purchase them would appreciate receiving them. Specifically, volumes 19 through 21 are missing from the library of Mieczyslaw Choynowski, Pracownia Psychologiczna, Szpital w Kobierzynie, Krakow 18, Poland.

Research Exchange

A Comment on the Wiener-Nichols Controversy (2, 3, 4, 5)

ROBERT R. HOLT
Research Center for Mental Health, New York University

To what a sad pass we have come when we must invoke their putative morale-building value as justification for projective techniques! Of course it is true that a professional person's habitual activities, and the accustomed heft of the tools of his trade, give him the reassurance of familiarity and renew his sense of professional identity and continuity with the past. But these are essentially non-rational considerations, and ones that can hardly play any major role in important decisions, such as whether projective techniques should be abandoned, should be subjected to intensified research, or the like. Doubtless the imposing beard of the Victorian physician was a great comfort to him, but that was not enough to preserve it when it was suspected of being a germ-carrier. Likewise, we have to find much more substantial merits in projective techniques than the non-rational ones cited by Weisskopf-Joelson and Nichols, to justify our keeping them.

This little controversy, which started out with a reasonable proposal by Wiener that we ought to work harder on getting people to agree on interpretations, seems to have been overexpanded by Nichols' rather extreme views. A column like this is hardly the place to debate fundamental issues about projective techniques or to make flat assertions like the two in Nichols' contribution.

Let's stick to the issue, because Wiener's point was a good one but was nearly lost in the hue and cry. Actually, we need to start further back than diagnoses, predictions or other integrative statements; there are entirely too few studies on the reliability of basic scoring. Having been involved in an attempt to devise scoring schemes that would guide clinical judgment in making predictions (1), I'm very much aware of how hard it is to get agreement on scores that require any degree of clinical judgement. Whether to persist through the laborious stages of developing and revising procedures for dealing with projective techniques or turn rather to more and more ingenious, objectively scorable tests is a decision that each of us has to make for himself in relation to a particular problem. For any specific research task, like predicting suitability of persons for an occupational role, it may be that the long-run usefulness and values of what can be developed will repay the effort; for routine diagnostic use, the situation is quite different. It is practically impossible to find objective instruments that will do much for us there, since we're not predicting a finite number of clearly specifiable criteria.

One of the ironic paradoxes in this situation is that the really research-minded people who might be willing to take up issues such as the one Wiener proposes are generally not much

interested in devoting their efforts to homely goals like improving the usefulness of clinical techniques. And, on the other hand, the good clinicians who are deeply interested in their instruments and how well they work often don't feel comfortable enough with research methods to do the job, either. Yet sooner or later we have to supplement the oral tradition of working with projective tests by research-derived data and principles.

If I may contribute a personal reflection to this discussion, I'd like to veer off in another direction and defend the proposition that projective techniques and other diagnostic tests are, and most usefully ought to remain, comparable to the interview and the autobiography as sources of clinical data. We hardly need better instruments for gathering data; rather, we need most of all a better conceptual framework in terms of which to conceptualize the richness of these data. Thus, even though I would favor studies of the kind Wiener advocates, I predict that in the long-run we won't make very much progress with projective techniques until we have some major advances in the theory of personality.

References

1. Holt, R. R. & Luborsky, L. *Personality patterns of psychiatrists.* New York: Basic Books, 1958. 2 vols.

2. Nichols, R. C. The problem of interjudge agreement and prediction: a contrary view. *J. proj. Tech.*, 1958, *22*, 255-257.

3. Weisskopf-Joelson, Edith. Comments on Nichol's and Wiener's notes on inter-judge agreement and prediction. *J. proj. Tech.*, 1959, *23*, 106.

4. Wiener, M. The problem of interjudge agreement and prediction. *J. proj. Tech.*, 1958, *22*, 122-124.

5. Wiener, M. A rejoinder to the problem of interjudge agreement and prediction. *J. proj. Tech.*, 1959, *23*, 105-106.